

**MINISTÉRIO DA DEFESA  
EXÉRCITO BRASILEIRO  
SECRETARIA DE CIÊNCIA E TECNOLOGIA  
INSTITUTO MILITAR DE ENGENHARIA  
CURSO DE MESTRADO EM QUIMICA**

**MARIA JEANEDITE DE SOUZA CARVALHO**

**QUIMIOMETRIA E ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO (NIR)  
E MEDIO (MIR) APLICADAS NA ANÁLISE DE ÓLEOS LUBRIFICANTES**

**Rio de Janeiro**

**2005**

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.



**INSTITUTO MILITAR DE ENGENHARIA**

**MARIA JEANEDITE DE SOUZA CARVALHO**

**QUIMIOMETRIA E ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO (NIR)  
E MEDIO (MIR) APLICADAS NA ANÁLISE DE ÓLEOS LUBRIFICANTES**

Dissertação de Mestrado apresentada ao Curso de  
Mestrado Química do Instituto Militar de Engenharia,  
como requisito parcial para a obtenção do título de Mestre  
em Ciências em Químicas.

Orientador: Prof. Luiz Eduardo Pizarro Borges - D. C.

Co-orientador: Flavio Santos de Gusmão Lima- D. C.

Rio de Janeiro

2005

C2005

INSTITUTO MILITAR DE ENGENHARIA

Praça General Tibúrcio, 80 – Praia Vermelha

Rio de Janeiro - RJ CEP: 22290-270

Este exemplar é de propriedade do Instituto Militar de Engenharia, que poderá incluí-lo em base de dados, armazenar em computador, microfilmar ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do(s) autor(es) e do(s) orientador(es).

543.08583 Carvalho, Maria Jeanedite de Souza  
C331q Quimiometria e Espectroscopia no Infravermelho próximo (NIR) e médio (MIR) aplicadas na análise de óleos lubrificantes / Maria Jeanedite de Souza Carvalho. - Rio de Janeiro: Instituto Militar de Engenharia, 2005. - 121 p.: il., graf., tab.

Dissertação (mestrado) – Instituto Militar de Engenharia – Rio de Janeiro, 2005.

1. Quimiometria. 2. Óleo Lubrificante. 3. Infravermelho. I. Título. II. Instituto Militar de Engenharia.

CDD 543.08583

**INSTITUTO MILITAR DE ENGENHARIA**

**MARIA JEANEDITE DE SOUZA CARVALHO**

**QUIMIOMETRIA E ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO (NIR)  
E MEDIO (MIR) APLICADAS NA ANÁLISE DE ÓLEOS LUBRIFICANTES**

Dissertação de Mestrado apresentada ao Curso de Mestrado em Química do Instituto Militar de Engenharia, como requisito parcial para a obtenção do título de Mestre em Ciências Químicas.

Orientador: Prof. Luiz Eduardo Pizarro Borges - D. C.

Co-orientador: Flavio Santos de Gusmão Lima – D.C.

Aprovada em 16 de Novembro de 2005 pela seguinte Banca Examinadora:

*Luiz Borges*

---

Prof. Luiz Eduardo Pizarro Borges – D. C. do IME - Presidente

*Flavio Santos de Gusmão Lima*

---

Flavio Santos de Gusmão Lima – D. C. da BR Distribuidora

*Alcino Palermo de Aguiar*

---

Prof. Alcino Palermo de Aguiar – D. C. do IME

*Fernando Cunha Peixoto*

---

Prof. Fernando Cunha Peixoto – D.C. do IME

Rio de Janeiro

2005

## AGRADECIMENTOS

Agradeço a todas as pessoas que me incentivaram, apoiaram e possibilitaram esta oportunidade de ampliar meus horizontes.

Aos meus familiares, cônjuge e mestres.

Em especial ao meu Professor Luiz Eduardo Pizarro Borges e ao Dr. Flavio Santos de Gusmão Lima, por suas disponibilidades e atenções.

A todos os meus amigos, que contribuíram direta ou indiretamente para realização do meu trabalho.

À CAPES/PROCAD pela bolsa concedida e pelos recursos para realização deste trabalho.



## SUMÁRIO

	LISTA DE ILUSTRAÇÕES.....	09
	LISTA DE TABELAS.....	11
	LISTA DE ABREVIATURAS E SÍMBOLOS.....	14
	LISTA DE SIGLAS.....	15
<b>1.0</b>	<b>INTRODUÇÃO</b> .....	<b>18</b>
1.1	OBJETIVOS.....	19
<b>2.0</b>	<b>LUBRIFICAÇÃO</b> .....	<b>20</b>
2.1	INTRODUÇÃO.....	20
2.2	LUBRIFICAÇÃO.....	22
2.3	ADITIVOS.....	23
2.4	PROPRIEDADES DOS ÓLEOS LUBRIFICANTES .....	25
2.4.1	ÍNDICE DE VISCOSIDADE OU IV.....	26
2.4.2	PONTO DE FLUIDEZ.....	26
2.4.3	VISCOSIDADE CINEMÁTICA.....	26
2.4.4	DENSIDADE .....	27
2.4.5	COR .....	27
2.4.6	ESPUMA .....	28
<b>3.0</b>	<b>ESPECTROMETRIA NO INFRAVERMELHO</b> .....	<b>29</b>
3.1	INTRODUÇÃO.....	29
3.2	FUNDAMENTAÇÃO TEÓRICA.....	30
3.2.1	ALTERAÇÃO NO MOMENTO DE DIPOLO.....	30
3.2.2	TIPOS DE VIBRAÇÃO MOLECULAR.....	30
3.2.3	MODELO DO OSCILADOR HARMÔNICO.....	31
3.2.4	MODELO QUÂNTICO DE UMA MOLÉCULA.....	33
3.2.5	BANDAS DE ABSORÇÃO NA REGIÃO DO INFRAVERMELHO .....	

	PRÓXIMO (NIR).....	35
3.2.6	BANDAS DE ABSORÇÃO NA REGIÃO DO INFRAVERMELHO MÉDIO (MIR).....	35
3.3	APLICAÇÕES.....	35
<b>4.0</b>	<b>QUIMIOMETRIA.....</b>	<b>37</b>
4.1	INTRODUÇÃO.....	37
4.2	PRÉ-TRATAMENTO DOS DADOS.....	39
4.2.1	DADOS CENTRADOS NA MÉDIA.....	39
4.2.2	ESCALAMENTO PELA VARIÂNCIA.....	40
4.2.3	AUTOESCALAMENTO.....	41
4.2.4	NORMALIZAÇÃO.....	42
4.2.5	DERIVAÇÃO.....	42
4.2.6	CORREÇÃO DE ESPALHAMENTO MULTIPLICATIVO (MSC).....	43
4.3	MÉTODOS DE ANÁLISE MULTIVARIADA.....	43
4.3.1	ANÁLISE EXPLORATÓRIA.....	44
4.3.1.2	ANÁLISE DE COMPONENTES PRINCIPAIS (PCA).....	44
4.3.1.2.a	ESCOLHA DO NÚMERO DE COMPONENTES PRINCIPAIS.....	46
4.3.1.2.b	GRÁFICOS NO PCA.....	47
4.3.2	MODELOS DE REGRESSÃO MULTIVARIADA.....	48
4.3.2.1	CALIBRAÇÃO E PREVISÃO MULTIVARIADA.....	48
4.3.2.2	ETAPA DA MODELAGEM.....	49
4.3.2.3	MÉTODOS DE CALIBRAÇÃO MULTIVARIADA.....	50
4.3.2.3.1	REGRESSÃO POR MÍNIMOS QUADRADOS PARCIAIS (PLS).....	50
4.3.2.4	MÉTODOS DE PREVISÃO MULTIVARIADA.....	53
4.3.2.4.1	MÉTODO DE ORDENAMENTO.....	54
4.3.2.4.2	VALIDAÇÃO CRUZADA.....	55
<b>5.0</b>	<b>METODOLOGIA.....</b>	<b>56</b>
5.1	OBTENÇÃO DAS AMOSTRAS.....	56
5.2	ANÁLISE DAS AMOSTRAS.....	59



5.3	CONJUNTO DE AMOSTRAS.....	59
5.4	LEITURA DOS ESPECTROS.....	60
5.4.1	ESPECTROS NIR.....	60
5.4.2	ESPECTROS MIR.....	60
5.5	TRATAMENTO DOS DADOS.....	62
5.5.1	DESENVOLVIMENTO DOS MODELOS.....	62
5.5.2	SELEÇÃO DAS VARIÁVEIS INDEPENDENTES.....	63
<b>6.0</b>	<b>RESULTADOS E DISCUSSÃO.....</b>	<b>66</b>
6.1	SELEÇÃO DAS AMOSTRAS E PROPRIEDADES.....	66
6.2	ESPECTROS DAS AMOSTRAS.....	66
6.3	CLASSIFICAÇÃO DAS AMOSTRAS.....	68
6.3.1	APLICAÇÃO DA PCA NOS DADOS DAS PROPRIEDADES.....	68
6.3.2	ANÁLISE DE PCA DOS ESPECTROS NIR .....	70
6.3.3	ANÁLISE DE PCA DOS ESPECTROS MIR .....	73
6.4	ESPECTRO COMPLETO.....	76
6.4.1	ESPECTRO COMPLETO – NIR.....	76
6.4.1.1	DISCUSSÃO– ESPECTRO COMPLETO (NIR).....	78
6.4.2	ESPECTRO COMPLETO – MIR.....	79
6.4.2.1	DISCUSSÃO – ESPECTRO COMPLETO (MIR) .....	82
6.5	ANÁLISE DE COMPONENTES PRINCIPAIS (PCA) .....	82
6.5.1	ANÁLISE DE COMPONENTES PRINCIPAIS – NIR.....	82
6.5.1.1	SELEÇÃO DE VARIÁVEIS - NIR .....	86
6.5.1.2	DISCUSSÃO – ANÁLISE DE COMPONENTES PRINCIPAIS (NIR).....	92
6.5.2	ANÁLISE DE COMPONENTES PRINCIPAIS – MIR.....	94
6.5.2.1	SELEÇÃO DE VARIÁVEIS - MIR.....	97
6.5.2.2	DISCUSSÃO – ANÁLISE DE COMPONENTES PRINCIPAIS (MIR).....	106
6.6	MÉTODO DE SELEÇÃO DE VARIÁVEIS $R^2$ .....	107
6.6.1	MÉTODO DE SELEÇÃO DE VARIÁVEIS $R^2$ – NIR.....	108
6.6.1.1	DISCUSSÃO - SELEÇÃO DE VARIÁVEIS $R^2$ (NIR).....	109
6.6.2	MÉTODO DE SELEÇÃO DE VARIÁVEIS $R^2$ – MIR.....	110
6.6.2.1	DISCUSSÃO - SELEÇÃO DE VARIÁVEIS $R^2$ (MIR) .....	113

6.7	RESUMO DOS MELHORES RESULTADOS.....	114
7.0	CONCLUSÕES.....	117
8.0	REFERÊNCIAS BIBLIOGRÁFICAS.....	119



## LISTA DE ILUSTRAÇÕES

FIG.2.1	Estrutura Molecular de óleos básicos .....	21
FIG. 3.1	Tipos de Vibrações Moleculares.....	31
FIG. 3.2	Modelo harmônico clássico para molécula diatômica.....	32
FIG. 3.3	Modelo anarmônico para molécula diatômica.....	34
FIG. 4.1	Conjunto de dados antes e depois de centrar na média.....	40
FIG. 4.2	Matrizes do PCA.....	45
FIG. 4.3	Matrizes do PLS.....	52
FIG. 5.1	Modelo de célula fixa de KBr usado no MIR.....	61
FIG. 5.2	Espectro de óleo lubrificante no MIR com caminho ótico 0,025 mm.....	62
FIG. 6.1	Espectro das amostras de óleos lubrificantes no NIR (2240-800 nm).....	67
FIG. 6.2	Espectro das amostras de óleos lubrificantes no MIR (4000-400 $\text{cm}^{-1}$ ).....	68
FIG. 6.3	Gráfico de <i>scores</i> PC <sub>1</sub> X PC <sub>2</sub> com todas as 71 amostras.....	69
FIG. 6.4	Gráfico de <i>scores</i> PC <sub>1</sub> X PC <sub>2</sub> dos espectros originais NIR das 71 amostras.....	71
FIG. 6.5	Gráfico de <i>scores</i> PC <sub>1</sub> X PC <sub>2</sub> dos espectros autoescalados NIR das 71 amostras.....	71
FIG. 6.6	Gráfico de <i>scores</i> PC <sub>1</sub> X PC <sub>2</sub> dos espectros normalizados NIR das 71 amostras.....	72
FIG. 6.7	Gráfico de <i>scores</i> PC <sub>1</sub> X PC <sub>2</sub> dos espectros normalizados e derivados NIR das 71 amostras.....	72
FIG. 6.8	Gráfico de <i>scores</i> PC <sub>1</sub> X PC <sub>2</sub> dos espectros originais MIR das 71 amostras.....	74
FIG. 6.9	Gráfico de <i>scores</i> PC <sub>1</sub> X PC <sub>2</sub> dos espectros autoescalados MIR das 71 amostras.....	74
FIG. 6.10	Gráfico de <i>scores</i> PC <sub>1</sub> X PC <sub>2</sub> dos espectros normalizados MIR das 71 amostras.....	75
FIG. 6.11	Gráfico de <i>scores</i> PC <sub>1</sub> X PC <sub>2</sub> dos espectros normalizados e derivados MIR das 71 amostras.....	75
FIG.6.12	Gráfico da Variância explicada versus Componentes principais (PCs) – NIR.....	83
FIG. 6.13	Gráfico dos Pesos versus PC 1 - NIR.....	84
FIG. 6.14	Gráfico dos Pesos versus PC 2 - NIR.....	84

FIG. 6.15	Gráfico dos Pesos versus PC 3 – NIR	85
FIG. 6.16	Gráfico do Resíduo versus <i>Leverage</i> – NIR	85
FIG. 6.17	Gráfico dos Pesos versus PC 1 – NIR: espectros normalizados	87
FIG. 6.18	Gráfico dos Pesos versus PC 2 – NIR: espectros normalizados	87
FIG. 6.19	Gráfico dos Pesos versus PC 3 – NIR: espectros normalizados	88
FIG. 6.20	Gráfico dos Pesos versus PC 1 – NIR: espectros normalizados e derivados	88
FIG. 6.21	Gráfico dos Pesos versus PC 2 – NIR: espectros normalizados e derivados	89
FIG. 6.22	Gráfico dos Pesos versus PC 3 – NIR: espectros normalizados e derivados	89
FIG. 6.23	Gráfico dos Pesos versus PC 1 – NIR: espectros autoescalados	90
FIG. 6.24	Gráfico dos Pesos versus PC 2 – NIR: espectros autoescalados	90
FIG. 6.25	Gráfico dos Pesos versus PC 3 – NIR: espectros autoescalados	91
FIG. 6.26	Gráfico de variância explicada versus componentes principais - MIR	94
FIG. 6.27	Gráfico dos Pesos versus componentes principais 1 – MIR	95
FIG. 6.28	Gráfico dos Pesos versus componentes principais 2 – MIR	95
FIG. 6.29	Gráfico dos Pesos versus componentes principais 3 – MIR	96
FIG. 6.30	Gráfico dos Pesos versus componentes principais 4 – MIR	96
FIG. 6.31	Gráfico do Resíduo versus <i>Leverage</i> – MIR	97
FIG. 6.32	Gráfico dos Pesos versus PC 1 – MIR: espectros normalizados	98
FIG. 6.33	Gráfico dos Pesos versus PC 2 – MIR: espectros normalizados	98
FIG. 6.34	Gráfico dos Pesos versus PC 3 – MIR: espectros normalizados	99
FIG. 6.35	Gráfico dos Pesos versus PC 4 – MIR: espectros normalizados	99
FIG. 6.36	Gráfico dos Pesos versus PC 1 – MIR: espectros normalizados e derivados	100
FIG. 6.37	Gráfico dos Pesos versus PC 2 – MIR: espectros normalizados e derivados	100
FIG. 6.38	Gráfico dos Pesos versus PC 3 – MIR: espectros normalizados e derivados	101
FIG. 6.39	Gráfico dos Pesos versus PC 4 – MIR: espectros normalizados e derivados	101
FIG. 6.40	Gráfico dos Pesos versus PC 1 – MIR: espectros autoescalamento	102
FIG. 6.41	Gráfico dos Pesos versus PC 2 – MIR: espectros Autoescalamento	102
FIG. 6.42	Gráfico dos Pesos versus PC 3 – MIR: espectros autoescalamento	103
FIG. 6.43	Gráfico dos Pesos versus PC 4 – MIR: espectros autoescalamento	103



## LISTA DE TABELAS

TAB. 2.1	Principais Aditivos de óleos Lubrificantes Acabados (LIMA, 2003).....	24
TAB. 3.1	Algumas aplicações na espectroscopia no infravermelho.....	36
TAB. 5.1	Produtos comerciais utilizados e a sua aplicação.....	56
TAB. 5.2	Propriedades dos óleos Minerais.....	57
TAB. 5.3	Propriedades dos óleos Semi-Sintéticos e Sintéticos.....	58
TAB. 5.4	Propriedades e Métodos de Referência – Óleos Lubrificantes.....	59
TAB. 5.5	Interpretação dos valores de $r^2$ .....	65
TAB. 6.1	Estatística dos dados – óleos Lubrificantes Comerciais.....	66
TAB. 6.2	PLS no NIR – Densidade. Espectro Completo. Validação: Validação Cruzada.....	77
TAB. 6.3	PLS no NIR- Viscosidade a 100 °C. Espectro Completo. Validação: Validação Cruzada.....	77
TAB. 6.4	PLS no NIR- Viscosidade a -15 °C. Espectro Completo. Validação: Validação Cruzada.....	77
TAB. 6.5	PLS no NIR- Ponto de Fluidez. Espectro Completo. Validação: Validação Cruzada.....	78
TAB. 6.6	PLS no NIR- Teor de Magnésio. Espectro Completo. Validação: Validação Cruzada.....	78
TAB. 6.7	Resultados da Validação no NIR. Espectros Completos. Validação: Validação Cruzada.....	78
TAB. 6.8	Amostras Seleccionadas para o Conjunto Teste. Método Ordenamento – Validação Externa.....	80
TAB. 6.9	PLS no MIR- Teor de Magnésio. Espectro Completo. Validação: Validação Cruzada.....	80
TAB. 6.10	PLS no MIR- Viscosidade a -15 °C. Espectro Completo. Validação: Validação Cruzada.....	80
TAB. 6.11	Modelo de Predição - Densidade. Espectro Completo. Seleção de amostras: Validação Externa.....	80
TAB. 6.12	Modelo de Predição - Ponto de Fluidez. Espectro Completo. Seleção de Amostras: Validação Externa.....	81
TAB. 6.13	Modelo de Predição - Viscosidade a 100°C. Espectro Completo. Seleção	

	de Amostras: Validação Externa.....	81
TAB. 6.14	Resultados da Validação no MIR. Espectros Completos. Validação : Validação Cruzada e Ordenamento.....	81
TAB. 6.15	PLS no NIR- Densidade. Seleção de Variáveis: PCA. Validação: Validação Cruzada.....	91
TAB. 6.16	PLS no NIR- Viscosidade a 100 °C. Seleção de Variáveis: PCA. Validação: Validação Cruzada.....	91
TAB. 6.17	PLS no NIR- Viscosidade a -15 °C. Seleção de Variáveis: PCA Validação: Validação Cruzada.....	92
TAB. 6.18	PLS no NIR- Ponto de Fluidez. Seleção de Variáveis: PCA. Validação: Validação Cruzada.....	92
TAB. 6.19	PLS no NIR- Teor de Magnésio. Seleção de Variáveis: PCA. Validação: Validação Cruzada.....	92
TAB. 6.20	Resultados da Validação no NIR. Seleção de variáveis: PCA. Validação : Validação Cruzada.....	92
TAB. 6.21	PLS no MIR- Teor de Magnésio. Seleção de Variáveis: PCA. Validação: Validação Cruzada.....	104
TAB. 6.22	PLS no MIR – Viscosidade a -15 °C. Seleção de variáveis: PCA. Validação: Validação Cruzada.....	104
TAB. 6.23	Modelo de Predição- Densidade. Seleção de variáveis: PCA. Seleção de Amostras: Validação Externa.....	104
TAB. 6.24	Modelo de Predição- Ponto de Fluidez. Seleção de variáveis: PCA. Seleção de Amostras: Validação Externa.....	105
TAB. 6.25	Modelo de Predição- Viscosidade a 100 °C. Seleção de variáveis: PCA. Seleção de Amostras: Validação Externa.....	105
TAB. 6.26	Resultados da Validação no MIR. Seleção de variáveis: PCA. Validação : Validação Cruzada e Ordenamento.....	105
TAB. 6.27	Número de onda retirado do modelo PLS – NIR.....	107
TAB. 6.28	Número de onda retirado do modelo PLS – MIR.....	107
TAB. 6.29	PLS no NIR- Densidade. Seleção de Variáveis: $r^2$ . Validação: Validação cruzada.....	108
TAB. 6.30	PLS no NIR- Viscosidade a 100 °C. Seleção de Variáveis: $r^2$ . Validação: Validação Cruzada.....	108



TAB. 6.31	PLS no NIR- Viscosidade a -15 °C. Seleção de Variáveis: $r^2$ . Validação: Validação Cruzada.....	108
TAB. 6.32	PLS no NIR- Ponto de Fluidez. Seleção de Variáveis: $r^2$ . Validação: Validação Cruzada.....	108
TAB. 6.33	PLS no NIR- Teor de Magnésio. Seleção de Variáveis: $r^2$ . Validação: Validação Cruzada.....	109
TAB. 6.34	Resultados da Validação no NIR. Seleção de Variáveis: $r^2$ . Validação : Validação Cruzada.....	109
TAB. 6.35	PLS no MIR- Teor de Magnésio. Seleção de Variáveis: $r^2$ . Validação: Validação Cruzada.....	111
TAB. 6.36	PLS no MIR- Viscosidade a -15 °C. Seleção de Variáveis: $r^2$ . Validação: Validação Cruzada.....	111
TAB. 6.37	Modelo de Predição- Densidade – Seleção de Variáveis: $r^2$ . Seleção de amostras: validação externa.....	111
TAB. 6.38	Modelo de Predição- Ponto de Fluidez – Seleção de Variáveis: $r^2$ Seleção de amostras: validação externa.....	112
TAB. 6.39	Modelo de Predição- Viscosidade a 100 °C – Seleção de Variáveis: $r^2$ Seleção de amostras: validação externa.....	112
TAB. 6.40	Resultados da Validação no MIR. Seleção de Variáveis: $r^2$ . Validação : Validação Cruzada e Ordenamento.....	112
TAB. 6.41	Resultados da Validação no NIR.....	114
TAB. 6.42	Resultados da Validação no MIR.....	114

## LISTA DE ABREVIATURAS E SÍMBOLOS

- A** – Absorbância
- a** – Constante para dada molécula
- d** – Distância
- E** – Matriz dos resíduos
- E** – Energia
- h** – Constante de Planck
- l** – Deslocamento dos átomos
- k** – Constante de força
- m** – Massa
- r** – Coeficiente de correlação
- R** – Matriz de correlação
- S** – Matriz da variância
- T** – Matriz dos escores
- X** – Matriz
- Y** – Matriz
- V** – Matriz dos pesos ou *loadings*
- x** – Coordenada linear na direção x
- x** – Constante de anarmonicidade
- y** – Coordenada linear na direção y
- v** – Número quântico vibracional
- μ** – Massa reduzida
- ν** – Freqüência vibracional da ligação
- t** – Transposta

## LISTA DE SIGLAS

NIR	Infravermelho próximo
MIR	Infravermelho médio
MLR	Regressão linear múltipla
PCA	Análise por componentes principais
PC	Componente principal
PCR	Regressão por componentes principais
PLS	Regressão por mínimos quadrados parciais
RMSEC	Raiz do quadrado do erro médio de calibração
RMSEP	Raiz do quadrado do erro médio de previsão
CLS	Mínimos quadrados clássico
ILS	Regressão inversa
ANP	Agência Nacional de Petróleo
ASTM	American Standard Methods
UNICAMP	Universidade Estadual de Campinas



## RESUMO

A aplicação das espectroscopias no infravermelho próximo (NIR) e infravermelho médio (MIR) foi investigada para a análise de óleos lubrificantes comerciais. Para isso, foram desenvolvidos modelos multivariados, correlacionando 5 propriedades físico-químicas (densidade, viscosidade a 100°C, ponto de fluidez, teor de magnésio e viscosidade a -15°C) com as absorvâncias no NIR (60 amostras de óleos lubrificantes) e no MIR (71 amostras de óleos lubrificantes). Os modelos foram obtidos pela técnica de regressão por mínimos quadrados parciais (PLS), mediante dois métodos de seleção de amostras (aleatório e ordenamento), quatro tipos de pré-tratamentos (sem pré-tratamento, autoescalamento, normalização e derivação após normalização), dois métodos de seleção de variáveis (PCA e matriz de correlação) e do emprego do espectro completo, sem seleção de variáveis.

As propriedades densidade e viscosidade a 100°C apresentaram modelos que só poderiam ser aplicados no controle do processo de produção de lubrificantes, pois os erros de predição apresentados são superiores a duas vezes os valores de reprodutibilidade dos respectivos métodos de referência. Para as propriedades teor de magnésio, ponto de fluidez e viscosidade a -15°C foram encontrados modelos satisfatórios que poderiam ser empregados em controle de qualidade e controle de processo.

Em relação às técnicas quimiométricas avaliadas, verificou-se que ambos os métodos de seleção de amostras fornecem modelos satisfatórios. Para ambas as técnicas espectroscópicas, os resultados foram satisfatórios, com ou sem método de seleção de variáveis, principalmente para os modelos de ponto de fluidez, teor de magnésio e viscosidade a -15°C. Apesar dos resultados com os espectros originais terem sido satisfatórios, em geral, os pré-tratamentos de primeira derivada dos espectros normalizados e de normalização dos espectros apresentaram os melhores resultados.



## ABSTRACT

035100617101

The application of near infrared (NIR) and mid infrared (MIR) spectroscopies for the analysis of commercial lubricant oil was investigated. Multivaried models had been developed, correlating 5 physical-chemical properties (density, viscosity at 100°C, pour point, magnesium content and viscosity at -15°C) with the absorbances in NIR (60 lubricant oil samples) and MIR (71 lubricant oil samples). The models were obtained using the partial least squares regression (PLSR) technique, using two methods for sample selection, four types of pre-treatments (without pre-treatment, scaling, normalization and derivation after normalization), two methods for variable selection and the use of the full spectra, without variable selection.

The models obtained for density and viscosity at 100°C can only be used in process control of lubricant production, as they presented prediction errors higher than two times the respective reference methods reproducibility values. For the others properties (pour point, magnesium content and viscosity at -15°C), the models were found satisfactory, and can be used in both quality control and process control.

Regarding the chemometric techniques, it was verified that both sample selection methods resulted in satisfactory models. The results were satisfactory for both spectroscopic techniques, with or without variable selection methods, mainly for the pour point, magnesium content and viscosity at -15°C models. Although the results for the original spectra models were satisfactory, in general, the first derivative of the normalized spectra and the normalization of the spectra pre-treatments have presented the best results.

## 1.0 INTRODUÇÃO

Preparados através da mistura de óleos básicos e aditivos, os óleos lubrificantes são produtos comerciais que precisam obedecer a vários parâmetros de qualidade. No Brasil, estes parâmetros são estabelecidos e controlados pela Agência Nacional de Petróleo (ANP); são propriedades físicas e químicas cujos valores precisam ser controlados durante sua produção.

Vários são os métodos analíticos conhecidos hoje na indústria, entretanto, é crescente a necessidade de métodos mais rápidos, precisos, de menor custo e que não poluam o meio ambiente.

Para tentar suprir esse tipo de necessidade, a espectroscopia no infravermelho tem sido muito pesquisada pelas indústrias de petróleo e petroquímica, devido à sua grande precisão, rapidez e eficiência, bem como à correlação entre os espectros dos derivados orgânicos e suas propriedades físico-químicas.

A técnica se destaca não só pela rapidez com a qual as amostras podem ser caracterizadas sem prévia manipulação, mas também pelo seu caráter não destrutivo e grande seletividade, permitindo determinações quantitativas de componentes em misturas complexas. Com o auxílio de técnicas quimiométricas, a espectroscopia de infravermelho pode ser usada no controle de produção e qualidade, reduzindo o tempo, a complexidade e o custo das análises.

A espectroscopia na região de infravermelho médio MIR ( $4000\text{ cm}^{-1} - 400\text{ cm}^{-1}$ ), apresenta as vibrações características (fundamentais) das ligações químicas, capazes de fornecer informações básicas sobre composição e estrutura de um composto.

Na espectroscopia da região do infravermelho próximo, NIR ( $4000\text{ cm}^{-1} - 12800\text{ cm}^{-1}$ ), observam-se bandas de absorção correspondentes a sobretons (harmônicas) das vibrações fundamentais e de combinações de vibrações C-H, O-H e N-H fundamentais. A informação analítica contida nas bandas sobrepostas do espectro NIR são dificilmente seletivas e são influenciadas por um número de variáveis estruturais, físicas e químicas. Além disso, diferenças entre as amostras podem causar também diferenças espectrais muito sutis, difíceis de serem distinguidas sem o auxílio de equipamentos e tratamentos apropriados.

A quimiometria consiste essencialmente na aplicação de técnicas matemáticas e estatísticas em problemas químicos. A técnica tem contribuído com a aplicação e divulgação de uma série de ferramentas matemáticas, que até então haviam sido largamente empregadas



principalmente em outras áreas como: engenharia, estatística, matemática e geologia (FILHO,2003).

A quimiometria relaciona variáveis analíticas com um número elevado de dados (como as absorvâncias nos diferentes comprimentos de onda em um espectro infravermelho) com uma (ou mais) propriedade das espécies químicas presentes na amostra (como a densidade ou viscosidade, por exemplo) através do emprego de técnicas de seleção de variáveis e regressão multivariada. O objetivo principal é construir um modelo que possa prever essa propriedade com precisão suficiente, tornando sua determinação mais fácil e rápida. É possível também identificar grupos de amostras com características similares por métodos de classificação (análise de *Clusters*).

## 1.1 OBJETIVOS

O presente trabalho tem como principal finalidade a elaboração de modelos de previsão para propriedades físico-químicas (densidade; viscosidade a 100°C e viscosidade a -15°C; ponto de fluidez e teor de magnésio) e a classificação de óleos lubrificantes automotivos comerciais através das espectroscopias de infravermelho médio e próximo associadas a técnicas quimiométricas.

Serão elaborados diferentes modelos baseados nos espectros de infravermelho médio e próximo utilizando-se diferentes métodos de tratamento dos dados originais.

A partir da seleção dos melhores modelos, espera-se sugerir uma técnica simples, rápida, segura, confiável e de baixo custo para a análise das propriedades citadas acima para óleos lubrificantes automotivos comerciais.

## 2.0 LUBRIFICAÇÃO

### 2.1 INTRODUÇÃO

Como arte, a lubrificação começou a ser praticada há milhares de anos. O estudo da lubrificação como uma ciência começou no século XVII, com o desenvolvimento dos rolamentos e dos eixos. A partir do século XIX, com o surgimento dos automóveis e dos motores de vapor, estimulou-se o desenvolvimento dos lubrificantes complexos modernos, formados por óleos básicos e aditivos químicos.

O estudo e o emprego de lubrificantes remonta à Idade Antiga, conforme nos demonstra a Arqueologia. Uma roda de oleiro da Mesopotâmia – 4000 A.C. – continha traços de betumem aderido ao seu mancal. Uma carruagem egípcia de 1400 A.C. possuía uma graxa primitiva (mistura de giz e gordura animal) no seu cubo de roda.

O primeiro pensador a abordar o fenômeno do atrito entre dois sólidos foi Aristóteles, ao redor de 350 A.C., segundo o seu método filosófico. O primeiro pesquisador a investigar tal fenômeno seguindo um método científico foi Leonardo da Vinci, em 1519, através de medidas das forças necessárias para movimentar blocos sobre superfícies horizontais e inclinadas. Ele descobriu que o atrito é uma força que se opõe ao movimento do bloco, proporcional ao peso do bloco e independente da área de contato entre os sólidos. Os blocos de superfícies polidas apresentaram força de atrito de aproximadamente um quarto do seu peso. Ele também descobriu que o emprego de pós finos e graxas na base dos blocos diminuía substancialmente a força de atrito.

O segundo pesquisador a investigar o fenômeno do atrito foi Guillaume Amontons, em 1699, através de experimentos que usavam molas para prover as forças para o movimento de objetos de madeira, cobre, ferro e chumbo lubrificados por graxa de origem animal.

O terceiro pesquisador a investigar tal fenômeno foi Euler, em 1750, que procurou medir as forças necessárias para movimentar blocos sobre superfícies inclinadas. Ele descobriu a diferença entre atrito estático e atrito dinâmico: o atrito é menor para os blocos que já se encontram em movimento.

O trabalho de Amontons foi estendido por Coulomb, em 1781, que empregou objetos dos mais diversos materiais e lubrificantes de diversos tipos. Ele acreditava que o atrito provinha



da integração entre as asperezas das superfícies sólidas. Em 1938, graças ao trabalho de Bowden e Tabor, foi demonstrado que a adesão (ligação) era um componente importante do atrito entre os sólidos.

Pesquisas mais profundas, abordando os tipos e os regimes de lubrificação, foram efetuadas por Petrov e Reynolds ao redor de 1883, por Hardy e Doubleday em 1922, e por diversos pesquisadores a partir de 1940. E finalmente, em 1965, a Sociedade Real dos Engenheiros de Lubrificação (Reino Unido) fundou formalmente a tribologia, que consiste na ciência que estuda as interações – atrito, desgaste e lubrificação – entre superfícies sólidas em movimento relativo.

A formulação dos lubrificantes é realizada através de uma combinação estudada e definida de óleos básicos e aditivos (misturados até a homogeneidade), de forma a garantir um nível de exigência adequado para o campo de utilização ao qual o lubrificante é destinado.

Os óleos minerais, obtidos do petróleo, são ainda os mais empregados em lubrificação, e têm suas propriedades relacionadas à natureza do óleo cru do qual foram originados e ao processo de refinação empregado. Na produção de lubrificantes, o petróleo é submetido inicialmente à destilação primária inicial ou topeamento (*topping*), onde as frações mais leves são removidas por destilação. Depois passa pela destilação a vácuo, separando-se diversas frações. A fração de óleos lubrificantes é submetida a tratamentos subseqüentes, tais como remoção de parafinas, remoção de asfalto, refinação ácida, refinação por solvente, etc. Os óleos lubrificantes minerais podem, então, ser classificados, de acordo com sua origem (naftênicos, parafínicos, aromáticos e, principalmente moléculas mistas) contendo geralmente quinze ou mais átomos de carbono por molécula (MOURA,1998).

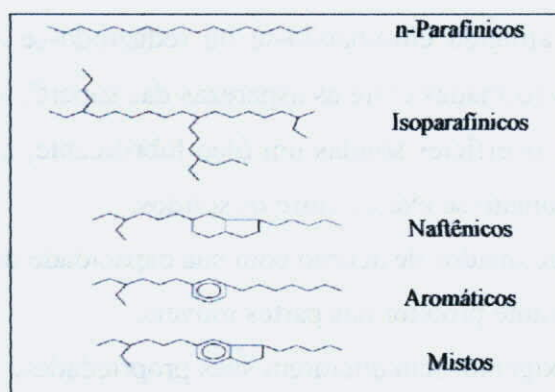


FIG. 1.1 Estrutura Molecular de óleos básicos

Existem também as bases sintéticas, que começaram a ser desenvolvidas no fim da década de 20 e ganharam impulso durante a II Guerra Mundial, principalmente na Alemanha, que buscava alternativas para os lubrificantes minerais, devido à escassez de petróleo. Além disso, no Pós-Guerra, novos tipos de motores e equipamentos, que exigiam lubrificantes com desempenho superior, passaram a ser fabricados em série.

A preocupação com a proteção do meio ambiente, que cresceu a partir da década de 50, impulsionou o uso e o aprimoramento dos lubrificantes sintéticos, tornando-os cada vez mais populares por possuírem Índices de Viscosidade mais elevados, menores pontos de fluidez, maior resistência à oxidação e menor volatilidade em relação aos óleos minerais.

São várias as bases sintéticas que podem ser empregadas na formulação dos lubrificantes. As de maior consumo são as "poli-alfa-olefinas" (PAOs), produzidas pela indústria petroquímica, seguindo-se os ésteres sintéticos e dos triglicerídeos naturais (óleo de soja, canola, mamona e girassol).

## 2.2 LUBRIFICAÇÃO

Lubrificar nada mais é do que aplicar uma substância entre duas superfícies em movimento relativo, formando uma película que evita o contato direto entre as superfícies, promovendo a redução do atrito, e, conseqüentemente, do desgaste e da geração de calor. A substância empregada com esta finalidade é o lubrificante, podendo ser um sólido, uma graxa ou um líquido.

A redução do atrito é atingida eliminando-se ou reduzindo-se a força necessária para separar (cisalhar) as junções formadas entre as asperezas das superfícies sólidas. Isto pode ser feito colocando-se entre as superfícies sólidas um óleo lubrificante, que sofrerá um aumento de pressão à medida que o contato se exerce entre os sólidos.

Os lubrificantes são selecionados de acordo com sua capacidade de:

- formar um filme deslizante protetor das partes móveis;
- resistir ao calor e ao oxigênio, sem alterarem suas propriedades;
- resistir a choques e cargas mecânicas sem alterar seu poder lubrificante;
- remover calor dos componentes internos do equipamento.



Para oferecer outras características de desempenho e proteção, são adicionados às bases lubrificantes alguns componentes químicos denominados aditivos.

### 2.3 ADITIVOS

Os aditivos são substâncias específicas adicionadas aos óleos lubrificantes com a finalidade de conferir ou realçar propriedades físicas. Na TAB. 1.1 estão descritos alguns dos principais aditivos e suas respectivas funções (LIMA, 2003).

TAB. 2.1 Principais Aditivos de óleos Lubrificantes Acabados (LIMA, 2003).

<b>Aditivos</b>	<b>Função</b>	<b>Substâncias Empregadas</b>
Anticorrosivos	Evitar a degradação da máquina por oxidação.	Dialquil ditiofosfato de zinco Sulfonatos metálicos alquilados
Antidesgaste	Reduzir o desgaste da maquina	Dialquil ditiofosfato de zinco Ácidos policarboxílicos Tricresil fosfato
Antiferrugem	Evitar a degradação das peças ferrosas	Fenóis alcoxilados Ácidos succínicos
Antifricção	Melhorar o desempenho dos motores	Glicerina substituída Compostos organometálicos
Extrema-pressão	Evitar a degradação da máquina	Compostos de enxofre e boro Compostos de boro e nitrogênio
Detergentes	Limpar a máquina	Sulfonatos metálicos alquilados Fenatos aromáticos ( $\phi$ -O-R-OH)
Dispersantes	Manter a sujeira em suspensão no óleo	Succinamidas polialquilênicas Polimetacrilatos
Antioxidantes	Evitar a degradação do óleo por oxidação	Dialquil diotiofosfato de zinco Aminas aromáticas 2,6-di-terc-butil-p-cresol
Antiespumantes	Evitar a formação de espuma no óleo	Polímeros de silicone
Passivadores de metais	Evitar a degradação do óleo por oxidação catalisada por metais	Tolutriazol Benzotriazol Tiodiazol
Biocidas	Evitar a degradação do óleo por microorganismos	Fenóis Azidas Benzoisotiazolinas
Melhoradores do índice de viscosidade	Reduzir a diminuição da viscosidade com o aumento da temperatura	Polimetacrilados Polibutenos Poliestirenos alquilados
Abaixadores do ponto de fluidez	Reduzir o ponto de fluidez do óleo	Polimetacrilatos Naftaleno alquilado Fenóis alquilados
Agentes de adesividade	Aumentar a adesividade do óleo as peças lubrificadas	Ésteres graxos
Demulgadores	Diminuir a formação de emulsão água/óleo	Poliéteres



## 2.4 PROPRIEDADES DOS ÓLEOS LUBRIFICANTES

As características de um lubrificante são divididas em 4 classes. Estas características podem ser determinadas através de ensaios laboratoriais padronizados.

- Características relacionadas com o escoamento – a viscosidade é quem define a espessura da película lubrificante no ponto lubrificado e por isso é considerada como a principal propriedade de um óleo lubrificante. A viscosidade pode ser medida em diferentes condições (temperatura e taxa de cisalhamento), dependendo da aplicação do lubrificante e de acordo com a especificação. Nas condições de trabalho, a viscosidade depende não só do índice de viscosidade (variação da viscosidade com a temperatura), mas também da ocorrência ou não de cristalização de parafinas (ARAUJO, 2000; FARAH, 1987);

- Características físico-químicas ligadas à composição (densidade, por exemplo), contaminação, deterioração ou falha de processamento – corrosividade ao cobre, índice de acidez, cinza, resíduo de carbono, ponto de fulgor e cor, ou ainda relacionadas com os aditivos empregados nas formulações;

- Composição – influenciam de forma determinante as propriedades físico-químicas do lubrificante. Envolve os tipos de hidrocarbonetos e de compostos contendo heteroátomos;

- Desempenho – pode ser determinado através de diversos tipos de ensaios: capacidade de proteção contra desgaste, estabilidade à oxidação, tendência à formação de espuma, demulsibilidade, etc.

Os ensaios laboratoriais mais utilizados para a especificação dos óleos lubrificantes comerciais estão descritos abaixo:

#### 2.4.1 ÍNDICE DE VISCOSIDADE OU IV

É a variação da viscosidade do lubrificante para uma determinada variação de temperatura. Durante a partida de máquinas lubrificadas (temperaturas baixas), deseja-se um rápido escoamento do óleo até as partes lubrificadas (baixa viscosidade) para facilitar o processo. A pleno funcionamento (temperaturas geralmente elevadas), deseja-se uma viscosidade relativamente alta, para manter a película lubrificante. Todavia, a viscosidade da maior parte dos líquidos decresce com o aumento da temperatura. Sendo assim, é desejável que esta variação seja a menor possível.

#### 2.4.2 PONTO DE FLUIDEZ

Trata-se da menor temperatura em que o óleo flui por ação da gravidade, depois de um resfriamento em condições padronizadas. Foi por muito tempo o único parâmetro utilizado para avaliar o comportamento dos óleos lubrificantes, quando submetidos a uma baixa temperatura. Tem grande importância, principalmente quando o produto formulado tem que atender a requisitos de viscosidade à baixa temperatura.

O resultado de seus ensaios depende do óleo básico e de sua resposta ao aditivo abaixador de ponto de fluidez empregados. Consiste no aquecimento preliminar de uma amostra, que é subsequentemente resfriada a uma taxa específica e observada em intervalos de 3°C. A menor temperatura na qual se observar movimento no óleo é anotada como o ponto de fluidez.

#### 2.4.3 VISCOSIDADE CINEMÁTICA

No que diz respeito à película lubrificante e à capacidade de suportar cargas, é a característica mais importante, pois representa a resistência que o fluido oferece ao escoamento ou a resistência interna oferecida pelas moléculas de um fluido quando expostas a uma tensão de cisalhamento.



É determinada pelo método ASTM D 445, que consiste nas seguintes etapas: colocação de um volume fixo de amostra num tubo viscosimétrico calibrado, no interior de um banho com temperatura específica e rigorosamente controlada. Em seguida, determina-se o tempo para que o volume da amostra escoe sob gravidade através de uma seção de um tubo de volume conhecido. A viscosidade é obtida através do produto do tempo de escoamento medido pela constante de calibração do tubo.

#### 2.4.4 DENSIDADE

É a relação entre a massa e o volume do óleo a uma determinada temperatura. Esta propriedade não é muito significativa para óleos lubrificantes, sendo empregada apenas para caracterizar os produtos e auxiliar na formulação dos mesmos. Para determinar a densidade de uma amostra, é introduzido um pequeno volume (0,7 mL aproximadamente) de amostra líquida em um tubo oscilante. Por fim, a variação na frequência de oscilação causada pela mudança na massa do tubo é usada em conjunto com dados de calibração (ARAUJO, 2000).

#### 2.4.5 COR

A inspeção visual dos óleos lubrificantes pode indicar a presença ou ausência de contaminantes grosseiros (outras fases líquida e/ou sólida) e o seu estado de oxidação (escurecimento anormal), além de servir de parâmetro para a padronização do produto.

A cor de um óleo básico é um sinal do grau de refino com o qual este foi fabricado, e é usada principalmente com o fim do controle da qualidade da produção do óleo lubrificante acabado. Uma grande alteração na cor de um óleo é sinal de contaminação. De qualquer modo, a cor não deve ser usada como único ensaio para a avaliação da qualidade de um produto. A cor do óleo é determinada por métodos comparativos: são empregados colorímetros SAYBOLT ou ASTM (método ASTM D1500: usando-se uma fonte padrão de luz, a amostra de um óleo é comparada com padrões – discos de vidro coloridos – aos quais se atribuíram valores de 0,5 à 8,0).

## 2.4.6 ESPUMA

Durante o funcionamento de um motor automotivo, as altas taxas de cisalhamento a que o óleo lubrificante é submetido na presença de ar podem acarretar a formação de espuma, estabilizada pelos aditivos surfactantes (detergentes, dispersantes, redutores de fricção e inibidores de corrosão) do óleo lubrificante. A presença de espuma num óleo lubrificante pode levar à deficiência de lubrificação e ao desgaste severo das peças do motor, além de acelerar a degradação do óleo básico.



### 3.0 ESPECTROMETRIA NO INFRAVERMELHO

#### 3.1 INTRODUÇÃO

Uma das primeiras aplicações da espectroscopia no infravermelho como ferramenta analítica ocorreu durante o período da II Guerra Mundial quando esta técnica foi utilizada no setor de controle de qualidade em algumas indústrias alemãs.

Uma vez que os espectros no infravermelho armazenam uma grande quantidade de informações sobre as amostras, eles apresentavam um elevado potencial para serem empregados nos mais diversos tipos de análises químicas e/ou físicas. Entretanto, até duas décadas atrás era praticamente impossível extrair informações quantitativas a partir dos espectros no infravermelho. Devido a este fato, a espectroscopia no infravermelho restringiu-se basicamente a aplicações qualitativas ou para reforçar hipóteses propostas sobre a estrutura química das espécies.

Em meados da década de oitenta, vários eventos nas áreas científicas e tecnológicas contribuíram para a inversão deste quadro, dentre eles, o desenvolvimento da microeletrônica e a popularização dos microcomputadores, que proporcionaram um significativo avanço nas análises instrumentais, possibilitando a aquisição de maneira fácil e rápida de um grande número de dados de uma mesma amostra. Dessa forma, o tratamento dos dados obtidos passou a exigir modelos mais complexos que iam além da tradicional calibração univariada. Com a aplicação de técnicas quimiométricas o problema da modelagem destes dados pôde ser solucionado. Assim, a quimiometria teve uma contribuição decisiva para a utilização da espectroscopia como uma ferramenta de análise em aplicações qualitativas e quantitativas na Química Analítica (FILHO, 2002).

O desenvolvimento e aperfeiçoamento das análises espectroscópicas no infravermelho também tiveram incentivos movidos pela crescente preocupação mundial com relação à questão ambiental. Estas análises, além de fornecerem resultados de maneira mais rápida, não são destrutivas e invasivas, assim como não geram subprodutos químicos tóxicos. Graças a estas vantagens, este tipo de análise tem sido aplicado no monitoramento em linhas de sistemas químicos industriais, no controle de qualidade, no combate à adulteração da composição química dos combustíveis, óleos comestíveis e alimentos, etc.

## 3.2 FUNDAMENTAÇÃO TEÓRICA

### 3.2.1 ALTERAÇÃO NO MOMENTO DE DIPOLO

A absorção de radiação eletromagnética por átomos ou moléculas exige que ela tenha energia apropriada e que haja um mecanismo de interação que permita a transferência de energia. O mecanismo apropriado à excitação vibracional é proporcionado pela variação periódica de dipolos elétricos na molécula durante as vibrações; a transferência de energia ocorre, então, por interação destes dipolos oscilatórios com o campo elétrico oscilatório da radiação infravermelha desde que a frequência com que ambos variam seja a mesma.

Por isso, para absorver radiação infravermelha, uma molécula precisa sofrer uma variação no momento dipolo como consequência do movimento vibracional ou rotacional. As moléculas diatômicas homonucleares como  $O_2$ ,  $N_2$  ou  $Cl_2$ , não apresentam nenhuma variação efetiva no momento dipolo e, assim, não absorvem na região do infravermelho (SKOOG, 2002).

### 3.2.2 TIPOS DE VIBRAÇÃO MOLECULAR

As posições relativas dos átomos em uma molécula não são exatamente fixas, variando constantemente em função dos diferentes tipos de vibrações e rotações referentes às ligações moleculares. Uma molécula contendo  $n$  átomos terá  $3n$  graus de liberdade. Para uma molécula não-linear, três graus de liberdade descrevem a rotação da molécula e três a translação. Os demais  $3n-6$  graus de liberdade correspondem aos graus de liberdade de vibração, ou vibração fundamental. As moléculas lineares possuem 1 modo normal a mais de vibração ( $3n-5$ ) (SILVERSTEIN, 1998).

Se tratando de moléculas diatômicas ou triatômicas, é relativamente simples definir o número e a natureza dessas vibrações e relacioná-las com as energias de absorção. Entretanto, em moléculas com elevado número de átomos, esta análise torna-se difícil, não só pela



quantidade de modos normais de vibração, mas também pelas múltiplas interações que podem ocorrer (ROCHA, 2002).

As vibrações podem ser divididas nas classes básicas de estiramento (*stretching*) e deformação angulares. Uma vibração de estiramento, envolve uma variação contínua na distância interatômica ao longo do eixo da ligação entre dois átomos. As vibrações de deformação angular são caracterizadas pela variação do ângulo entre duas ligações, podendo ser divididas em quatro tipos: vibração de deformação angular simétrica no plano, tipo “tesoura” (*scissoring*), balanço (*rocking*), sacudida (*wagging*) e torção (*twisting*). Os vários tipos de ligação estão mostrados na FIG. 3.1 (SKOOG, 2002).

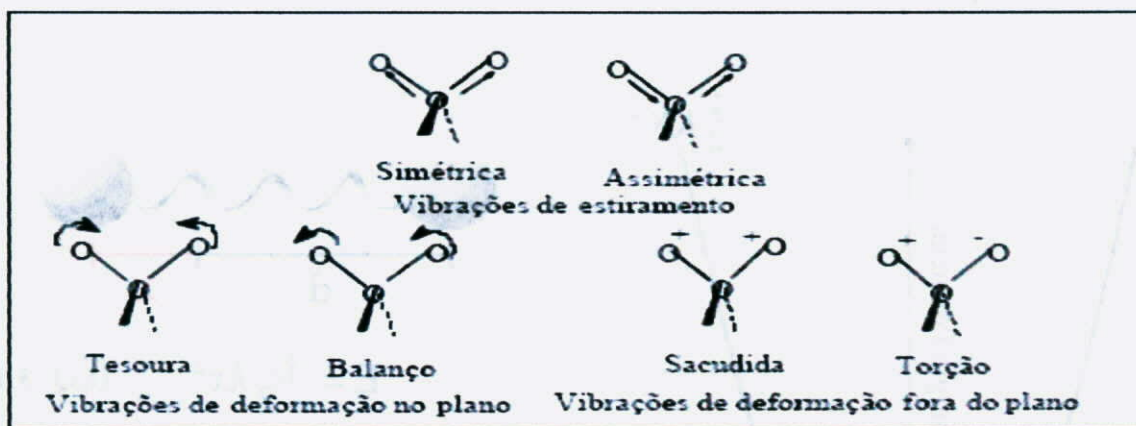


FIG. 3.1 Tipos de Vibrações Moleculares

### 3.2.3 MODELO DO OSCILADOR HARMÔNICO

A radiação no infravermelho, absorvida por uma molécula, faz as ligações individuais destas vibrarem com maior amplitude e frequência, de maneira similar ao que ocorre num oscilador diatômico.

O conceito básico da espectroscopia vibracional se baseia no comportamento de uma molécula diatômica, semelhante a um oscilador harmônico. Sendo assim, as moléculas ativas no infravermelho podem ser vistas como componentes de modelos mecânicos simples. Para uma molécula diatômica, o modelo pode ser descrito como uma mola de constante  $k$  presa a duas bolas de massas  $m$  (FIG. 3.2).

Quando a mola está na posição  $d$ , possui energia potencial zero. Se a mola for comprimida ou esticada de uma distância  $x$ , a energia potencial  $E$  aumenta de acordo com a EQ. 3.1. A parábola formada, representada na FIG. 3.2, é chamada de curva de potencial harmônico.

A frequência de vibração  $\nu$  da molécula relaciona-se com a constante  $k$  e massa  $m$  pela EQ. 3.2, de maneira que quanto menor a massa do átomo e maior a constante, mais alta será a frequência resultante.

Neste modelo, quando a mola é esticada até uma distância máxima arbitrária  $x$  e solta, ela vibra com energia total igual a  $\frac{1}{2}kx^2$ . Portanto, o modelo possibilita que a molécula vibre com qualquer energia total do sistema. Apesar de ser correto para uma mola e duas esferas, esta é uma afirmação incorreta para uma molécula (ROCHA, 2002).

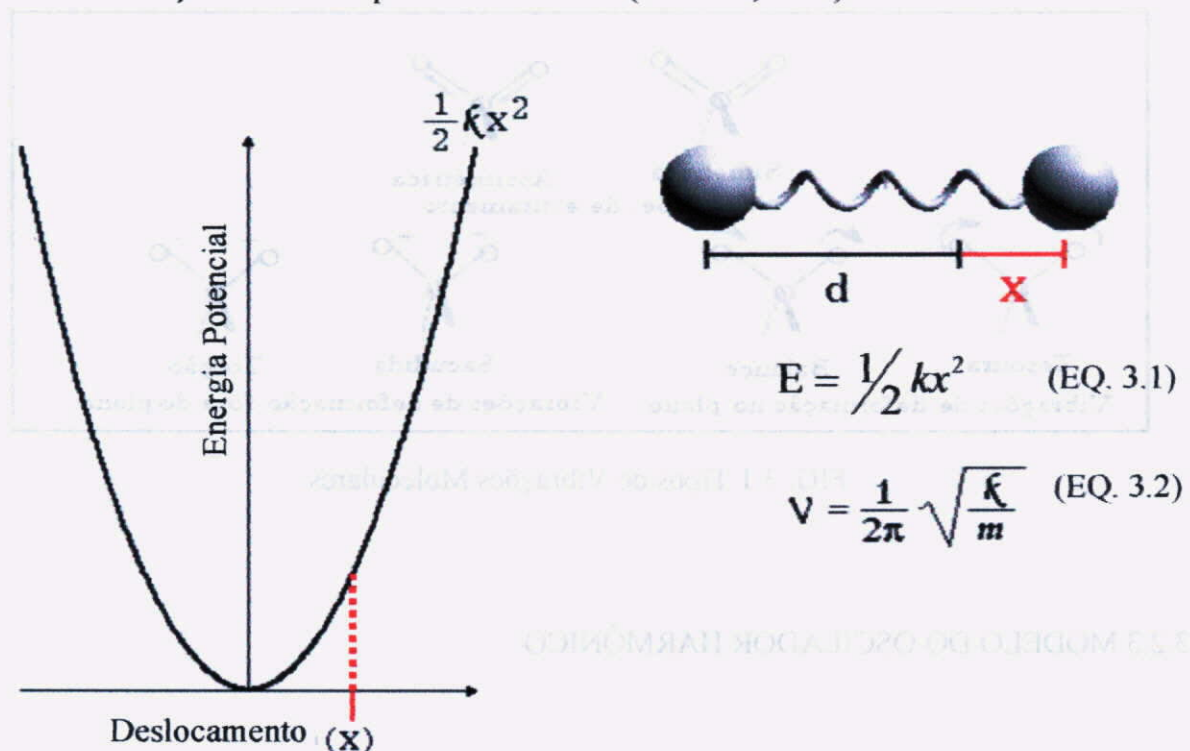


FIG. 3.2 Modelo harmônico clássico para molécula diatômica.



### 3.2.4 MODELO QUÂNTICO DE UMA MOLÉCULA

O modelo quântico prevê basicamente as mesmas relações entre frequência de vibração, constante de força e massa que são previstas pelo modelo clássico descrito no item anterior. Entretanto, a mecânica quântica diz que uma molécula somente pode vibrar com níveis de energia  $E_n$ , dados pela EQ. 3.3, onde  $h$  é a constante de Planck. Desta maneira, a energia é dita como quantizada.

$$E_n = \left( n + \frac{1}{2} \right) h\nu, \text{ onde } n = 0, 1, 2, 3.. \quad (\text{EQ. 3.3})$$

Pela mecânica quântica, a molécula também só pode emitir ou absorver energia igual à diferença entre dois níveis. Para um oscilador harmônico, essa transição só pode ocorrer entre um nível e o outro imediatamente próximo, isto é,  $\Delta n = \pm 1$ . É a chamada regra de seleção.

Pela regra de seleção, uma molécula somente pode absorver luz com energia igual a  $h\nu$ . Desta forma, o espectro de infravermelho de uma molécula diatômica conteria um pico único na frequência correspondente a esta energia. Essa transição ocorre na chamada frequência fundamental.

Entretanto, uma molécula real não é verdadeiramente um oscilador harmônico. À medida que dois átomos se aproximam, a repulsão colombiana entre os dois núcleos produz uma força que age na mesma direção da força de restauração da ligação; assim, espera-se que a energia potencial cresça mais rapidamente do que a prevista no modelo do oscilador harmônico. No outro extremo da oscilação, um decréscimo na força de restauração e, portanto, da energia potencial, ocorre quando a distância interatômica se aproxima daquela em que ocorre a dissociação dos átomos. Este comportamento pode ser resolvido com um modelo de potencial anarmônico.

Embora o modelo harmônico possa ajudar a compreender a espectroscopia vibracional, ele possui algumas restrições para a espectroscopia no infravermelho próximo (NIR), porque não permite transições onde  $\Delta n$  seja maior que 1. Assim, faz-se necessário a construção de um modelo mais realista, onde possam existir transições com  $\Delta n = \pm 2$  ou  $\Delta n = \pm 3$ , chamadas de sobretons (*overtones*) ou harmônicas (não-fundamentais), em frequências aproximadamente duas a três vezes a da linha fundamental ( $2h\nu$  e  $3h\nu$ ); a intensidade de

absorção dessas linhas harmônicas é frequentemente pequena e os picos podem não ser observados (FIG. 3.3).

Os espectros vibracionais são ainda mais complicados, pelo fato de que duas vibrações diferentes em uma molécula podem interagir para dar picos de absorção com frequências que são aproximadamente as somas ou as diferenças das suas frequências fundamentais. Novamente, as intensidades dos picos de combinação ou diferença geralmente são baixas (SKOOG, 2002).

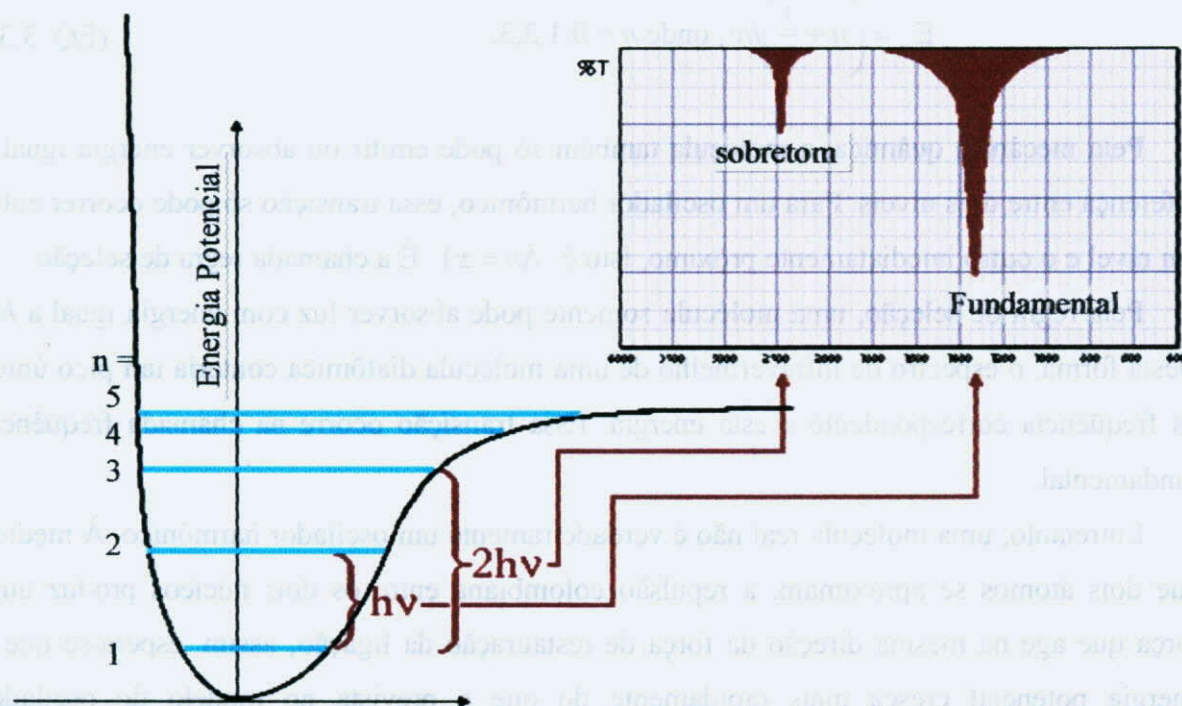


FIG. 3.3 Modelo anarmônico para molécula diatômica.

Entre deformações, a interação ocorre na presença de ligações comuns aos grupos vibracionais. O acoplamento entre vibrações de estiramento e deformação ocorre caso a ligação de estiramento forme um dos lados do ângulo da deformação. Praticamente não existe interação entre átomos afastados por duas ou mais ligações. O acoplamento de vibrações é um fenômeno comum; como resultado, a posição de um pico de absorção correspondente a um dado grupo pode variar bastante no espectro.



### 3.2.5 BANDAS DE ABSORÇÃO NA REGIÃO DO INFRAVERMELHO PRÓXIMO (NIR)

A intensidade de uma dada banda de absorção está associada com a magnitude da variação do dipolo durante o deslocamento de átomos em uma vibração e com seus graus de anarmonicidade. Ambos os fenômenos estão presentes com alta intensidade em ligações envolvendo hidrogênio e alguns outros elementos, tais como carbono, nitrogênio, enxofre e oxigênio. As ligações O-H, C-H, N-H e S-H tendem a apresentar alta anarmonicidade e alta energia de ligação com transições vibracionais fundamentais na região de 3000 nm - 4000 nm. Conseqüentemente, pode-se prever que os sobretons e combinações de vibrações fundamentais de tais ligações devem ocorrer na região de energia associada com fótons NIR. As intensidades são cerca de 10 vezes menores para combinação e mais do que 1000 vezes menores para sucessivos sobretons, com relação à absorção resultante das vibrações fundamentais (ALBUQUERQUE, 2004).

### 3.2.6 BANDAS DE ABSORÇÃO NA REGIÃO DO INFRAVERMELHO MÉDIO (MIR)

A espectroscopia na região de infravermelho médio MIR (4000 nm – 400 nm), é com certeza a região mais utilizada, pois apresenta as vibrações características (fundamentais) das ligações químicas, sendo empregada tanto para análise qualitativa, onde é capaz de fornecer informações sobre as características estruturais das moléculas, como para análise quantitativa, devido à correlação entre os espectros dos derivados orgânicos e suas propriedades físico-químicas.

## 3.3 APLICAÇÕES

A espectroscopia de infravermelho NIR e MIR inclui várias aplicações em diversas áreas. Alguns exemplos estão descritos na TAB.3.1.



TAB. 3.1 Algumas Aplicações na Espectroscopia no Infravermelho

Técnica	Aplicação	Referência
MIR	Aplicação de algoritmos genéticos para determinação simultânea de maltose, glicose e frutose.	FILHO, 2002
NIR	Avaliar as propriedades de óleo diesel de refinarias.	SIMÃO, 1999
NIR e MIR	Comparar as técnicas na determinação da curva de destilação do querosene.	CHUNG, 1999
NIR	Controle de processos, particularmente em aplicações <i>on-line</i> e <i>at-line</i> em indústrias petroquímicas.	BOKOBZA, 1998.
NIR	Controle do processo produtivo de óleos lubrificantes básicos.	LIMA, 2005
MIR	Desenvolver um método para classificação do tipo e do grau de uso do lubrificante	BORIN, 2003
NIR e MIR	Desenvolvimento de modelos matemáticos para a previsão da densidade, T50, T90, MON e RON da gasolina.	ROCHA, 2002
NIR	Determinação do teor de princípio ativo na detecção de medicamentos falsificados.	SCAFI, 2000
NIR	Determinação de BTEX (Benzeno, tolueno, etilbenzeno e xilenos), que são os constituintes mais solúveis da gasolina, presentes em amostras aquosas.	ALBUQUERQUE, 2004
NIR	Determinação dos teores de cinzas e proteínas em farinha de trigo.	FERRÃO, 2004
NIR	Determinação de proteínas em café cru.	MORGANO, 2005
MIR	Determinação quantitativa do teor de sacarose em açúcar cristal.	MORGANO, 2003
NIR	Determinação simultânea e não-destrutiva de sólidos totais e solúveis, licopeno e beta-caroteno em produtos de tomate.	PEDRO, 2004
NIR	Predizer o potencial carcinogênico de óleos lubrificantes básicos através da determinação total das concentrações de poliaromáticos.	LIMA, 2003
MIR	Previsão do MON em gasolina.	ANDRADE, 1999
NIR	Quantificar as duas contaminações mais importantes dos óleos lubrificantes (gasolina e etilenoglicol).	PASHOAL, 2003

## 4.0 QUIMIOMETRIA

### 4.1 INTRODUÇÃO

A quimiometria é um ramo da química que visa, entre outras coisas, extrair o máximo possível de informações de dados químicos, a partir de métodos estatísticos e matemáticos. Foi em 1972 que o químico sueco Svante Wold formalizou uma área de estudo de aplicação de métodos matemáticos às ciências químicas, propondo originalmente o termo quimiometria (*chemometrics*).

Na quimiometria, existe uma classificação dos métodos analíticos baseada nas respostas que fornecem. Existem os métodos de ordem zero, onde uma amostra fornece apenas uma variável independente (a resposta desta amostra é um escalar); nesta classe estão os métodos clássicos (ou via úmida) e os métodos eletroquímicos. Os métodos de primeira ordem são os métodos onde a análise de uma amostra gera diversas variáveis independentes, ou seja, a resposta desta amostra é um vetor. Nesta classe estão os métodos espectroscópicos e cromatográficos. Existem ainda os métodos de segunda ordem, em que uma amostra gera duas dimensões de variáveis independentes, de forma que a resposta desta amostra é uma matriz; nesta classe se incluem os métodos acoplados: cromatografia gasosa/espectrometria de massas, espectrometria de massa/espectroscopia no infravermelho e ressonância magnética nuclear bidimensional, entre outros.

Os métodos de calibração também podem ser divididos em classes, assim como os métodos analíticos. Entretanto, a classe de um método de calibração possui sempre uma ordem superior à ordem do método analítico para o qual foi gerado, de forma que a calibração de um instrumento de ordem zero é uma calibração de primeira ordem, já que uma coleção de escalares gera um vetor. No caso de instrumentos de primeira ordem, a calibração é de segunda ordem, pois um conjunto de vetores determina uma matriz. No caso dos métodos de segunda ordem, a calibração passa a ser de terceira ordem, trabalhando-se com tensores (LIMA, 2003).

Seguindo esta classificação, observamos que as espectroscopias no NIR e no MIR podem ser classificadas como métodos de análise instrumental de primeira ordem, uma vez que um espectro fornece um vetor com várias coordenadas, de acordo com a quantidade de



comprimentos de onda deste espectro. Sendo assim, exigem calibrações de segunda ordem que irão gerar uma matriz.

Como exemplo pode-se citar o que foi utilizado no presente trabalho: obteve-se um conjunto de calibração com 71 amostras lidas na região do MIR (primeira ordem), com o equipamento ajustado para intervalos de  $2 \text{ cm}^{-1}$ , que forneceu 794 leituras espectrais para cada amostra. Obteve-se, então, uma matriz  $X_{71 \times 794}$  (segunda ordem) de variáveis independentes, onde cada linha representa uma amostra.

$$X = \begin{bmatrix} X_{11} & \dots & X_{1794} \\ \vdots & \dots & \vdots \\ X_{711} & \dots & X_{71794} \end{bmatrix}$$

A cada uma das amostras pode-se relacionar o resultado de um ou mais testes primários (por exemplo, viscosidade a  $100^\circ\text{C}$ ). Estes resultados formariam um vetor  $y_{71}$  de variáveis dependentes, para uma propriedade, ou uma matriz  $Y_{71 \times 5}$ , caso sejam analisadas 5 propriedades simultaneamente.

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_{71} \end{bmatrix} \dots \text{ou} \dots Y = \begin{bmatrix} Y_{11} & \dots & Y_{15} \\ \vdots & \dots & \vdots \\ Y_{711} & \dots & Y_{715} \end{bmatrix}$$

Não é difícil perceber a necessidade de computadores relativamente rápidos e de métodos estatísticos específicos para que a correlação entre essa grande quantidade de dados possa ser obtida.

O método dos mínimos quadrados clássico (CLS) mostra-se inadequado para esses casos. O CLS requer o conhecimento prévio de todas as substâncias envolvidas na análise, que, no caso específico de frações de petróleo, não é possível.

Atualmente, os métodos de regressão dos componentes principais (PCR) e de mínimos quadrados parciais (PLS) são as alternativas mais utilizadas. São os mais usuais porque utilizam toda a informação contida no espectro, lidam bem com colinearidade, ruídos e suportam um bom grau de não-linearidade do modelo. Além disso, não sofrem alteração sensível nos parâmetros com a inclusão de novas amostras nos conjuntos de calibração (métodos robustos).

O PCR e o PLS baseiam-se na análise de componentes principais (PCA) que é uma técnica linear de redução de dimensionalidade da matriz de dados. Esta técnica identifica as direções ortogonais de variância máxima dos dados originais e reconstrói esses valores em um espaço de menor dimensão formada do subconjunto das componentes de mais alta variância.



## 4.2 PRÉ-TRATAMENTO DOS DADOS

Segundo (BEEBE, 1999), antes de se iniciar qualquer tipo de análise multivariada, é preciso realizar uma manipulação matemática prévia do conjunto de dados. Normalmente, nas etapas de manipulação, preparação, condicionamento ou aquisição dos dados, ocorrem variações nos procedimentos experimentais que causam variações espectrais nos dados amostrais e podem dar origem a amostras anômalas (*outliers*). Para eliminar ou reduzir o impacto dessas variáveis, sem afetar as variações espectrais fundamentais relacionadas com a propriedade a ser modelada, realizam-se pré-tratamentos matemáticos específicos para cada fonte de variação.

As principais fontes de variações espectrais que se deseja eliminar são: intensidade do ruído, interação entre os componentes da mistura, variação no percurso óptico, efeitos da fonte e do detector, deslocamentos de comprimento de onda, desvios da linha de base e efeito do espalhamento da radiação.

Vale salientar que, apesar das vantagens, todos os métodos matemáticos usados podem denegrir a informação espectral esperada, geralmente através da diminuição da razão sinal/ruído. De forma que, a otimização apropriada dos parâmetros que tais métodos requerem deve ser investigada. Serão descritas abaixo algumas técnicas de pré-tratamento capazes de eliminar ou reduzir variações espectrais aleatórias ou não-uniformes.

### 4.2.1 DADOS CENTRADOS NA MÉDIA

Esta técnica é normalmente utilizada na construção de modelos de calibração, e é empregada quando há presença de sinais caracterizados como ruído. Obtém-se o valor médio para cada variável (EQ. 4.1) que é subtraído de cada um dos termos da variável correspondente (EQ. 4.2). Desta forma, o sistema de coordenadas é deslocado para o centro dos dados, ou seja, o eixo de coordenadas translada para um centro, de modo a colocar os dados numa forma mais conveniente à análise e visualização. Este procedimento é aplicado às colunas da matriz (FIG. 4.1).

$$\bar{X}_j = \frac{1}{n} \sum_{i=1}^n X_{ij} \quad (\text{EQ. 4.1})$$

$$X_{ij(cm)} = X_{ij} - \bar{X}_j \quad (\text{EQ. 4.2})$$

Onde:

$X_{ij}$  = Valor da variável j na amostra i;

$\bar{X}_j$  = Média dos valores das amostras na coluna j;

$X_{ij(cm)}$  = Valor centrado na média para a variável j na amostra i;

n = Número de amostras.

Na FIG. 4.1 pode-se observar que a direção do vetor  $\mu_1$  acompanha a maior variação dos dados (pontos). O vetor  $\mu_1$  representa, no caso, a primeira componente de um PCA.

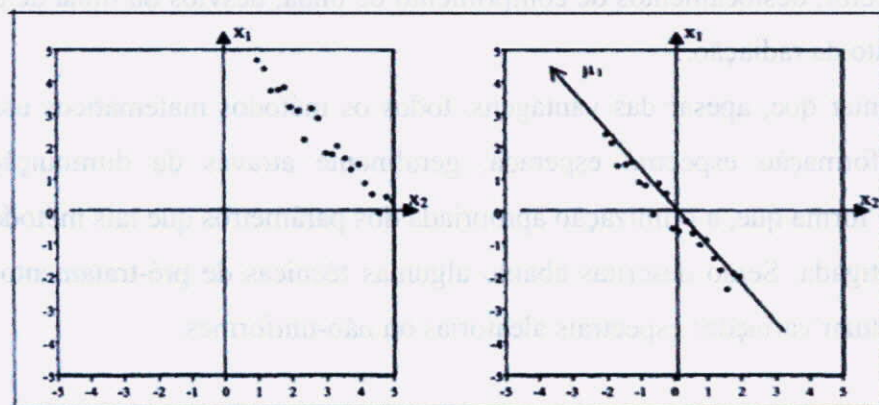


FIG. 4.1 conjunto de dados antes e depois de centrar na média

#### 4.2.2 ESCALAMENTO PELA VARIÂNCIA

Este tipo de pré-tratamento é utilizado quando as variáveis possuem dimensões muito discrepantes entre si, sendo necessário considerarem o peso das variáveis em diferentes escalas de forma equivalente. Isso é feito através da divisão de cada valor da variável j pelo respectivo desvio padrão, como na EQ. 4.3. Dessa forma, minimiza-se o risco de perder informações relevantes.

Assim,

$$X_{ij(ev)} = \frac{X_{ij}}{S_j} \quad (\text{EQ. 4.3})$$

Em que:

$X_{ij(ev)}$  = Valor escalado pela variância para a variância  $j$  na amostra  $i$ ;

$X_{ij}$  = Valor da variável  $j$  na amostra  $i$ ;

$S_j$  = desvio padrão dos valores da variável  $j$ , calculado a partir da variância  $S^2$  dada pela EQ. 4.4.

$$S_j^2 = \frac{1}{n-1} \sum_{i=1}^n (X_{ij} - \bar{X}_j)^2 \quad (\text{EQ. 4.4})$$

Onde:

$n$  = número de amostras;

$X_{ij}$  = valor da variável  $j$  na amostra  $i$ ;

$\bar{X}_j$  = média dos valores das amostras na coluna  $j$ , calculado pela EQ. 4.1

#### 4.2.3 AUTOESCALAMENTO

O autoescalamento aplica ambas as técnicas descritas acima de uma só vez, de modo que a transformação realizada sobre o conjunto original dos dados resulta em variáveis com média zero e variância igual a um (autoescalamento). Desta forma será dada a mesma importância para todas as variáveis, independente da sua dimensão. (EQ. 4.5)



$$X_{ij(as)} = \frac{X_{ij} - \bar{X}_j}{S_j} \quad (\text{EQ. 4.5})$$

Onde:

$X_{ij(as)}$  = valor autoescalado da variável j para a amostra i;

$\bar{X}_j$  = média dos valores das amostras na coluna j, calculado pela EQ. 4.1

$S_j$  = desvio padrão dos valores da variável j, calculado a partir da variância  $S^2$  dada pela EQ. 4.4.

#### 4.2.4 NORMALIZAÇÃO

A normalização é realizada através da divisão de todo o espectro por um valor adequado – uma constante. De forma geral, é através do emprego do valor máximo, médio ou mínimo da intensidade apresentada por um comprimento de onda do espectro, que os dados de cada espectro são escalados. Com isso, a área ou comprimento do vetor espectral torna-se normalizado (BEEBE, 1999).

Efeitos espectrais multiplicativos, como variações no percurso óptico, efeito da fonte e do detector e efeitos de espalhamento das ondas eletromagnéticas, podem ser corrigidos através da normalização de um espectro no NIR e no MIR, sendo esse o seu principal objetivo.

#### 4.2.5 DERIVAÇÃO

O objetivo da derivação de um espectro no NIR é aumentar a resolução espectral a partir da eliminação dos desvios da linha base, permitindo a identificação de bandas fracas ou sobrepostas, visivelmente identificáveis no espectro original. Vale salientar que, para a derivação de um espectro, é necessário fazer um balanço entre o aumento da resolução do

espectro e a diminuição da razão sinal/ruído, que é a principal desvantagem deste pré-tratamento.

Para poder calcular a derivada “suavizada” de um espectro, outro artifício é necessário, já que este é representado por uma seqüência de pontos (uma variável discreta, e não uma função). Pode-se usar a técnica de Savitsky-Golay, através do uso de coeficientes polinomiais (BEEBE, 1999).

#### 4.2.6 CORREÇÃO DE ESPALHAMENTO MULTIPLICATIVO (MSC)

O método de correção de espalhamento multiplicativo é um pré-tratamento específico para espectros de infravermelho próximo, e é normalmente aplicado na correção de linha base. Esta técnica garante que a informação química seja preservada e que as diferenças entre os espectros sejam minimizadas.

Este método assume que os comprimentos de onda da luz espalhada possuem uma dependência distinta entre a luz espalhada e absorvida pelos constituintes da amostra. Portanto, teoricamente é possível separar estes dois sinais. Através deste método, tenta-se remover o efeito do espalhamento pela linearização de cada espectro por um espectro ideal. Para efeito de cálculo, considera-se que o espectro ideal é o espectro médio do conjunto de dados para o qual deseja realizar a correção da linha base. Em seguida, calcula-se o coeficiente angular e linear do gráfico entre o espectro ideal e o espectro que vai ser corrigido, através de uma regressão linear. O espectro corrigido é calculado subtraindo cada ponto do espectro do valor do coeficiente linear e dividindo este valor pelo coeficiente angular (FILHO, 2003).

#### 4.3 MÉTODOS DE ANÁLISE MULTIVARIADA

São vários os tipos de técnicas de estatística multivariada, que estão associadas às mais variadas aplicações. A utilização de um método ou outro, ou mesmo a combinação deles, depende da natureza do problema que se enfrenta, ou do tipo de informação que se deseja obter por meio de análises multivariadas. Esses métodos podem ser classificados em dois conjuntos principais: métodos de análise exploratória e modelos de regressão multivariada.



#### 4.3.1 ANÁLISE EXPLORATÓRIA

Este método é realizado antes de se iniciar o tratamento do conjunto de dados, e tem como objetivo a avaliação inicial dos dados para verificar quais informações podem ser extraídas. Assim, é possível definir diretrizes para um tratamento mais aprofundado. Isto é feito através da utilização de algoritmos que permitem reduzir a dimensão dos dados, ou organizá-los numa estrutura que facilite a visualização de todo o conjunto de forma global. Para realizar esta análise inicial, os dois métodos estatísticos mais conhecidos são a Análise de Componentes Principais (PCA) e a Análise Hierárquica de Agrupamento (HCA). Neste trabalho, utilizou-se a PCA, que será abordada mais detalhadamente.

##### 4.3.1.1 ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

Por volta da década de 60, a análise por componentes principais (PCA), ainda com o nome de análise de fatores principais, foi utilizada pela primeira vez por Malinowki para o tratamento de dados químicos (ALBUQUERQUE, 2004). Poucos anos depois foram publicados inúmeros trabalhos sobre diferentes aplicações deste método na resolução de problemas químicos.

A análise por componentes principais (PCA) é uma técnica que consiste em projetar o máximo de informação no menor número possível de eixos não correlacionados, funcionando como um método de compressão da informação. Esses eixos são combinações lineares das variáveis originais, com pesos escolhidos de forma a maximizar a variância, tomada como uma medida da informação. Assim, a primeira componente principal (PC1) representa a direção de máxima variância (máxima informação) no espaço multidimensional em que o conjunto de dados está representado. A segunda componente (PC2), que é ortogonal à primeira, descreve o máximo da informação ainda não descrita por PC1, e assim por diante. Dessa forma, cada amostra é representada por um ponto no espaço de dimensão  $M$ , do qual se podem extrair informações das amostras e dos agrupamentos de amostras que apresentam características semelhantes.



No exemplo  $X_{71 \times 794}$  (seção 4.1) o PCA resume toda a informação contida na matriz  $X$  e a coloca em forma passível de ser compreendida, reduzindo-se o espaço  $R_{794}$  em um espaço menor  $R_3$  ou  $R_4$ , por exemplo.

O PCA funciona transformando a matriz de dados originais  $X_{(n \times m)}$ , formada por  $n$  objetos e  $m$  variáveis, no produto de duas variáveis menores, conforme mostra a EQ. 4.6,

$$X_{(n \times m)} = T_{(n \times k)} V'_{(k \times m)} + E_{(n \times m)} \quad (\text{EQ. 4.6})$$

Em que:

$T$  é a matriz dos escores e contém as novas coordenadas das amostras no novo sistema de eixo das PCs. Cada amostra é descrita pela sua projeção nas componentes principais, no lugar das variáveis originais.

$V$  é a matriz dos pesos ou *loadings* que é composta dos autovetores associados aos autovalores de maior módulo da matriz de variâncias e covariâncias dos dados originais.

A informação que não está contida nestas matrizes sobra como uma “variância inexplicada de  $X$ ” na matriz dos resíduos ( $E$ ) com mesmas dimensões de  $X$ .

A FIG. 4.2 apresenta essas matrizes graficamente, onde  $k$  é o número de componentes principais incluídas na análise. Como  $k$  normalmente é menor que  $m$  e  $n$ , uma matriz maior é transformada em duas matrizes menores, tornando a análise da matriz  $X$  mais simples (ALBUQUERQUE, 2004).

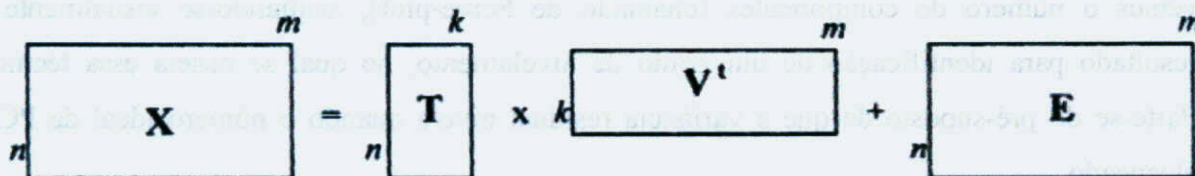


FIG. 4.2 Matrizes do PCA

#### 4.3.1.1.a ESCOLHA DO NÚMERO DE COMPONENTES PRINCIPAIS

Em um modelo, o número de PCs pode variar bastante, sendo que, do ponto de vista prático, quanto maior esse número maior a dificuldade na interpretação dos resultados. Para contornar isso, na análise de componentes principais, uma das etapas mais importantes é decidir quantas PCs devem ser extraídas (número de componentes significantes ou dimensionalidade do modelo), de forma que a cada nova PC extraída aumenta-se a quantidade de informação (variância) explicada pelo modelo.

Vários métodos podem ser utilizados na seleção do número ideal de PCs do modelo, como, por exemplo, as seguintes técnicas:

**Porcentagem de variância explicada:** Para se obter 100% de variância explicada, basta usar todas as componentes disponíveis. Entretanto, deve-se eliminar parte desta variância na construção do modelo, já que é composta de erro aleatório (ruído). Isto pode ser feito pela comparação da variância remanescente com o nível de ruído esperado nas medições (BEEBE, 1999), de forma que o valor de 90% é freqüentemente encontrado na literatura. De acordo com OTTO (1997), este critério é basicamente heurístico e pode ser usado quando existe experiência suficiente na análise de conjunto de dados similares.

**Scree-test:** É feita através da construção de um gráfico de variância ou de autovalores versus o número de componentes (chamado de Scree-plot), avaliando-se visualmente o resultado para identificação de um ponto de nivelamento, no qual se baseia essa técnica. Parte-se do pré-suposto de que a variância residual nivela quando o número ideal de PC é alcançado.

**Validação cruzada:** Normalmente retiram-se as amostras, uma a uma (*leave-one-out*), e calcula-se o modelo com os  $k$  componentes principais das amostras restantes. Repete-se o procedimento para todas as  $n$  amostras e calcula-se o resíduo.

Constrói-se, então, um gráfico de erro de previsão (PRESS ou RMSEC) versus PCs, anotando-se o mínimo da curva para a escolha do número ideal de componentes. Não havendo um mínimo, pode-se avaliar o ponto de nivelamento visualmente, assim como no



scree-test. Um gráfico de comportamento errático pode sugerir a existência de amostras anômalas.

#### 4.3.1.1.b GRÁFICOS NO PCA

A análise de PCA também é um meio de simplificar a série de dados. Com o emprego dos fatores (PCs) é possível utilizar gráficos com uma, duas ou três dimensões a fim de identificar variáveis que contêm informações similares ou informações completamente independentes. A maneira mais prática de se obter essas informações é através de análise dos gráficos das matrizes relacionadas com o PCA (V,T e E). Segue abaixo uma descrição desses gráficos.

**Gráfico dos scores (T)** - Pode revelar agrupamentos (*clusters*) ou tendências das amostras analisadas, geralmente de difícil visualização no caso das variáveis originais. Pode, ainda, destacar amostras cujo comportamento seja muito diferente do restante do conjunto (anômalas). Além disso, a oposição das amostras no gráfico pode servir para interpretar as componentes principais e, nesse contexto cada PC pode ser interpretada como um resumo das características distintas de cada amostra.

**Gráfico dos loadings (V)** – O gráfico de loadings permite testar a homogeneidade das contribuições das variáveis de X no modelo, representando as variáveis originais das componentes principais no espaço (em 2D ou 3D). O loading de cada variável indica o quanto essa variável participa na definição da PC (o quadrado do loading fornece sua porcentagem na PC), de forma que as variáveis com baixa influência no modelo são representadas por loadings pequenos (em módulo).

**Gráfico dos Resíduos versus leverage** - A *leverage* é uma medida da influência de uma amostra no modelo. Amostras com *leverage* grande influenciam fortemente o modelo e, normalmente, podem ser vistas como afastadas do agrupamento principal no gráfico de *scores*. Geometricamente, a *leverage* pode ser interpretada como a distância de uma amostra ao centróide ( $X_1, X_2, \dots, X_n$ ) do conjunto de dados e pode ser calculada pela EQ. 4.7 através da

distância da amostra ao centro do conjunto, verificando-se a distância Euclidiana no espaço das componentes principais.

$$h = T * T^T \quad (\text{EQ. 4.7})$$

Onde

h representa o valor da *leverage*

T é a matriz dos escores

T<sup>T</sup> é a transposta de T

Vale salientar que nem sempre *leverages* altos na construção do modelo representam bons resultados, uma vez que amostras anômalas, por exemplo, costumam apresentar medidas elevadas de *leverage*, sendo capazes de forçar o modelo a descrever características que podem não refletir o perfil do restante do conjunto de dados. No entanto isso não significa que amostras com alto valor de *leverage* devem ser excluídas, cada caso deve ser avaliado cautelosamente.

#### 4.3.2 MODELOS DE REGRESSÃO MULTIVARIADA

Os métodos de regressão multivariada têm por objetivo a modelagem dos dados medidos independentemente, com algumas propriedades das amostras. As aplicações mais freqüentes dos métodos de regressão são em calibração e previsão multivariada.

##### 4.3.2.1 CALIBRAÇÃO E PREVISÃO MULTIVARIADA

A calibração é o processo de construção de um modelo matemático que relaciona a resposta de um instrumento a certas propriedades das amostras. Por exemplo, na determinação de um composto químico em diferentes amostras, muitas vezes o instrumento utilizado para



se determinar o teor do composto não fornece uma resposta direta. É necessária a construção de um modelo de regressão que correlacione a resposta instrumental ao teor analito, ou outra informação qualquer que se deseje obter. Este modelo poderá então ser utilizado para prever a concentração do analito em amostras desconhecidas, processo conhecido como previsão (RIBEIRO, 2003).

Quando existe apenas uma resposta instrumental por amostra, por exemplo valor de absorvância para o analito que está sendo investigado, os métodos de calibração e previsão são ditos univariados, isto é, dependem apenas de uma variável que será relacionada em termos matemáticos ao teor de apenas um analito.

A calibração e a previsão multivariadas são processos utilizados quando mais de uma resposta instrumental são utilizadas, como por exemplo, nos métodos espectrais, como análises de infravermelho médio e próximo. Neste caso, o instrumento fornece mais de uma resposta por amostra e também é possível analisar mais de um analito simultaneamente, o que não seria possível utilizando um método univariado (RIBEIRO, 2003).

#### 4.3.2.2 ETAPA DA MODELAGEM

A modelagem é a construção do modelo matemático que melhor descreva a correlação entre a matriz X (variáveis independentes) e o vetor y (variáveis dependentes). Primeiramente, é feita a seleção do conjunto de amostras a ser utilizado. Isto porque amostras com comportamento ou características muito distintas do restante do conjunto podem interferir de forma negativa e, dependendo do caso, devem ser excluídas.

Após escolher as amostras que melhor represente os dados que se deseja modelar e efetuar o pré-tratamento, a próxima etapa é a construção do modelo matemático, que pode ser encontrado através do vetor de regressão  $\beta$  como representado pela EQ. 4.8.

$$y = X * \beta \quad (\text{EQ. 4.8})$$

A construção deste modelo pode ser feita por vários métodos. A escolha do método mais apropriado está relacionada ao perfil do conjunto de dados. Alguns dos métodos mais comuns

para a construção de modelos de regressão multivariada serão descritos nos métodos de calibração multivariada.

#### 4.3.2.3 MÉTODOS DE CALIBRAÇÃO MULTIVARIADA

A construção dos modelos de regressão multivariada pode ser feita a partir de vários métodos, sendo que a escolha de cada um deles está relacionada ao perfil do conjunto de dados. Basicamente, os métodos de regressão multivariados podem ser classificados em dois tipos principais: os métodos dos mínimos quadrados clássicos (CLS) e os métodos dos mínimos quadrados inversos (ILS), sendo este último tipo composto pelos métodos de regressão linear múltipla (MLR), regressão por componentes principais (PCR) e mínimos quadrados parciais (PLS).

Neste trabalho, utilizou-se o método dos mínimos quadrados parciais, PLS, que será abordada mais detalhadamente.

Os demais métodos de regressão não foram utilizados no presente trabalho, mas pode-se encontrar informações detalhadas sobre os mesmos em diversos textos, como BEEBE(1999), OTTO (1999) e MONTGOMERY, 2003.

##### 4.3.2.3.1 REGRESSÃO POR MÍNIMOS QUADRADOS PARCIAIS (PLS)

Neste método de calibração multivariado geram-se, simultaneamente, dois conjuntos de variáveis: o primeiro descreve as amostras com relação as variáveis independentes originais e o segundo descreve as amostras com relação à variável dependente.

Procede-se da seguinte forma: para o primeiro conjunto, realiza-se um PCA com as variáveis independentes originais, e com a variável dependente. Ou seja, aplica-se a PCA indiretamente na matriz dos dados  $\mathbf{X}$   $n \times m$ , através da matriz ampliada  $\mathbf{X}'$ , que possui dimensão  $n \times (m+1)$ . Da mesma forma, e paralelamente, o segundo conjunto de componentes principais é obtido empregando-se as variáveis independentes originais em conjunto com as



variáveis dependentes. Os processos podem ser descritos pelas equações abaixo e representados pela FIG.4.4.

$$X = T * V^T + E = \sum_{i=1}^k t_i v_i^t + E \quad (\text{EQ. 4.9})$$

Onde:

X = matriz de dados originais

T = matriz dos escores de X

$V^T$  = matriz transposta dos pesos de X ou matriz dos *loadings* de X

E = matriz dos resíduos de X

t = coluna de T para i-ésima PC

$v_i^t$  = pesos (*loadings*) para a i-ésima PC

i = cada componente principal

k = numero de componentes principais utilizado na modelagem

$$Y = U * Q^T + F = \sum_{i=1}^k u_i q_i^t + F \quad (\text{EQ. 4.10})$$

Onde:

Y = matriz de dados originais

U = matriz dos escores de Y

$Q^T$  = matriz transposta dos pesos de Y ou matriz dos *loadings* de Y

F = matriz de resíduos de Y

u = coluna de U da i-ésima PC

$q_i^t$  = pesos (*loadings*) para a i-ésima pC

i = cada componente principal

k = número de componentes utilizados na modelagem

$$\begin{array}{c}
 \begin{array}{c}
 \boxed{\text{X}} \begin{array}{l} m \\ \hline \end{array} \\
 \begin{array}{l} n \\ \hline \end{array}
 \end{array}
 =
 \begin{array}{c}
 \boxed{\text{T}} \begin{array}{l} k \\ \hline \end{array} \\
 \begin{array}{l} n \\ \hline \end{array}
 \end{array}
 \times
 \begin{array}{c}
 \boxed{\text{V}^t} \begin{array}{l} m \\ \hline \end{array} \\
 \begin{array}{l} k \\ \hline \end{array}
 \end{array}
 +
 \begin{array}{c}
 \boxed{\text{E}} \begin{array}{l} m \\ \hline \end{array} \\
 \begin{array}{l} n \\ \hline \end{array}
 \end{array}
 \\
 \\
 \begin{array}{c}
 \boxed{\text{Y}} \begin{array}{l} z \\ \hline \end{array} \\
 \begin{array}{l} n \\ \hline \end{array}
 \end{array}
 =
 \begin{array}{c}
 \boxed{\text{U}} \begin{array}{l} k \\ \hline \end{array} \\
 \begin{array}{l} n \\ \hline \end{array}
 \end{array}
 \times
 \begin{array}{c}
 \boxed{\text{Q}^t} \begin{array}{l} z \\ \hline \end{array} \\
 \begin{array}{l} k \\ \hline \end{array}
 \end{array}
 +
 \begin{array}{c}
 \boxed{\text{F}} \begin{array}{l} z \\ \hline \end{array} \\
 \begin{array}{l} n \\ \hline \end{array}
 \end{array}
 \end{array}$$

FIG. 4.3 Matrizes do PLS

A correlação linear entre os escores do bloco Y com os escores do bloco X, chamadas de relações internas, pode ser feita através da EQ. 4.11 para cada componente de cada vez, fornecendo o modelo mais simples para esta correlação.

$$u_i = \beta * t_i \quad (\text{EQ. 4.11})$$

Onde o coeficiente  $\beta$  pode ser descrito pela EQ. 4.12

$$\beta = \frac{u_i^t t_i}{t_i^t t_i} \quad (\text{EQ. 4.12})$$

Calculando-se os *loadings* da matriz X, temos (EQ. 4.13).

$$P^T = \frac{t^T X}{t^T t} \quad (\text{EQ. 4.13})$$

Para as matrizes X e Y, formam-se novos resíduos, e a matriz final do método PLS pode ser escrita como:

$$Y = \beta T \cdot Q + F \quad (\text{EQ. 4.14})$$



Uma das limitações deste modelo é que os escores são calculados separadamente para X e Y, sendo construída a relação entre eles somente depois, o que pode resultar num modelo insatisfatório, como, por exemplo, numa relação não linear entre X e Y.

Existe um tratamento alternativo que permite calcular os escores dos dois blocos de forma a produzir a melhor relação linear entre eles. Isto é feito fazendo-se uma rotação dos eixos das componentes principais. Como os escores são as projeções dos pontos nestes eixos, conseqüentemente seus valores são mudados de maneira a permitir que a melhor relação linear entre os escores de X e de Y seja construída.

Tanto o PCR quanto o PLS, desenvolvem modelos separados para cada variável dependente, entretanto, o cálculo das variáveis latentes é realizado de forma diferente: enquanto que no PCR aplica-se uma PCA na matriz X, no PLS duas PCA's são realizadas, nas matrizes X e Y, usando-se todas as variáveis originais disponíveis em cada uma delas. Sendo que, para o PLS, as primeiras variáveis latentes são, geralmente, suficientes para descrever satisfatoriamente a variação encontrada no modelo.

#### 4.3.2.4 MÉTODOS DE PREVISÃO MULTIVARIADA

Uma vez construído o modelo, é necessário que ele seja testado, ou validado para garantir que os valores das variáveis dependentes estimados sejam iguais, ou muito próximos dos valores obtidos experimentalmente.

Neste caso, um grupo de amostras de propriedade conhecidas (conjunto de validação) é separado do conjunto original e o modelo é construído com as amostras restantes (conjunto de modelagem). Este modelo é então utilizado para prever as propriedades das amostras deste conjunto de validação. Para obter uma boa estimativa da habilidade de previsão, o conjunto de validação precisa ser representativo.

Neste trabalho, a habilidade de previsão dos modelos de calibração para o PLS no infravermelho foi estimada pelo método de ordenamento e pelo método de validação cruzada.

#### 4.3.2.4.1 MÉTODO DE ORDENAMENTO

O método de ordenamento pertence à estratégia geral de empregar os valores da propriedade físico-químicos do grupo inicial de amostras visando à seleção de algumas destas para o conjunto de validação (BEEBE, 1999).

O método consiste em ordenar as amostras por ordem decrescente dos valores das propriedades para o qual se deseja desenvolver o modelo. Uma vez ordenadas as amostras, divide-se a lista obtida em tantos grupos quanto forem as amostras do conjunto validação desejado. Finalmente, retira-se o mais simetricamente possível de cada um destes grupos uma amostra para o conjunto validação; as amostras restantes são utilizadas no conjunto de modelagem.

As propriedades previstas são então comparadas com as propriedades obtidas experimentalmente e o erro de previsão é calculado (EQ. 4.15). A eficiência do modelo, ou seja, sua capacidade preditiva, é avaliada de acordo com a dimensão deste erro. Se o erro for pequeno, o modelo pode ser utilizado com confiança para prever as propriedades desconhecidas de outras amostras. Se este erro for grande significa que o modelo não é eficaz.

$$\text{erro} = (\hat{y}_i - y_i) \quad (\text{EQ. 4.15})$$

Onde

$\hat{y}_i$  é o valor previsto para a amostra  $i$  utilizando o modelo;

$y_i$  é o valor medido para a amostra

Em seguida, calcula-se a soma dos quadrados dos erros de previsão (PRESS), dada pela EQ. 4.16, ou ainda a raiz quadrada dos erros de previsão (RMSEP), dada pela EQ. 4.17.

$$PRESS = \sum_i (\hat{y}_i - y_i)^2 \quad (\text{EQ. 4.16})$$



$$RMSEP = \sqrt{\frac{\sum_i (\hat{y}_i - y_i)^2}{n}} \quad (\text{EQ. 4.17})$$

Onde

n = número de amostras do conjunto de calibração

#### 4.3.2.4.2 VALIDAÇÃO CRUZADA

Primeiramente, retira-se um pequeno conjunto de amostras, constrói-se o modelo com as amostras restantes e faz-se a previsão para aquelas amostras. A seguir, as amostras retiradas são incluídas novamente no conjunto de modelagem, outras são retiradas e se repete o processo até que todas as amostras do conjunto de modelagem sejam testadas (FERREIRA, 1999).

A validação cruzada pode ser feita retirando-se apenas uma amostra de cada vez, dependendo do conjunto de dados que está sendo estudado, da quantidade de amostras, do número de amostras de cada subconjunto, etc.

Os valores previstos são então comparados com os valores obtidos experimentalmente e o erro de previsão é calculado, como descrito anteriormente pela EQ. 4.15. Em seguida, calcula-se a soma dos quadrados dos erros de previsão (PRESS), dada pela EQ. 4.16, ou ainda a raiz quadrada dos erros de previsão (RMSEP), dada pela EQ. 4.17.

A validação do modelo permite testar sua eficiência, ou seja, se o modelo é realmente capaz de prever as propriedades (variáveis dependentes) a partir das variáveis independentes de forma confiável. Uma vez validado o modelo, ele poderá ser utilizado para fazer a previsão de propriedades em amostras desconhecidas.

## 5.0 METODOLOGIA

### 5.1 OBTENÇÃO DAS AMOSTRAS

Neste trabalho foram utilizadas 71 amostras de óleos lubrificantes comerciais, minerais e sintéticos da BR e da BARD AHL (TAB. 5.1). As amostras foram adquiridas (BR) ou cedidas (BARD AHL) e as empresas informaram gentilmente os valores das propriedades para os diferentes óleos e lotes (TAB.s 5.2 e 5.3).

Para manter a confidencialidade dos lubrificantes, por solicitação das indústrias, os nomes das amostras foram substituídos por símbolos.

As amostras foram colocadas em vidros, etiquetadas e armazenadas em local seco e ao abrigo da luz no laboratório de processos do IME, sendo retiradas apenas para as leituras dos espectros no infravermelho médio e próximo.

TAB. 5.1 Produtos comerciais utilizados e sua aplicação

Lubrificantes	Tipo	Aplicação
A1	Mineral	Monograu utilizado em motores a gasolina, a álcool e a automóveis adaptados para uso de gás natural.
A2	Mineral	Multiviscoso utilizado em motores a gasolina e álcool
A3	Mineral	Multiviscoso utilizado em motores a gasolina e álcool
A4	Mineral	Multiviscoso utilizado em motores a gasolina e álcool
A5	Semi-Sintético	Multigrau proteger o motor contra as duras condições de trafego urbano encontradas nos grandes centros
A6	Sintético	Multiuso utilizado em motores a gasolina e álcool em qualquer tipo de serviço severo
B1	Mineral	Multiviscoso de ultima geração para motores a álcool, gasolina e GNV.
B2	Sintético	Multiviscoso de alta tecnologia contem óleos sintéticos e aditivos que atendem o nível de desempenho API SL.



TAB. 5.2 Propriedades dos óleos Minerais

Lubrificante Mineral	Densidade g/cm <sup>3</sup>	Viscosidade		Ponto de Fluidez°C	Teor de Mg%(m)
		100°CcSt	-15°C cP		
1	0,8962	14,36	-	-15	-
2	0,8963	14,68	-	-12	-
3	0,896	14,64	-	-9	-
4	0,8924	14,60	-	-6	-
5	0,8918	14,57	-	-6	-
6	0,8922	14,40	-	-15	-
7	0,8963	14,80	-	-18	-
8	0,8906	14,44	-	-9	-
9	0,8929	14,42	-	-9	-
10	0,8954	14,40	-	-9	-
11	0,8943	14,68	-	-6	-
12	0,8888	14,20	7700	-24	0,102
13	0,8903	14,83	8300	-24	0,106
14	0,8908	14,75	8300	-24	0,103
15	0,8899	14,69	4950	-23	0,100
16	0,8892	15,03	7750	-27	0,095
17	0,8898	13,83	8583	-27	0,097
18	0,8902	14,91	6400	-24	0,100
19	0,8908	13,98	8700	-27	0,104
20	0,8875	20,49	8026	-30	0,090
21	0,8903	20,77	8800	-27	0,097
22	0,8892	21,20	8233	-24	0,092
23	0,8877	19,04	7375	-24	0,073
24	0,8925	17,21	8700	-24	0,085
25	0,8921	18,98	8900	-24	0,098
26	0,8886	18,53	8402	-27	0,086
27	0,8925	18,65	8600	-24	0,088
28	0,8926	18,65	8600	-27	0,091
29	0,8910	18,89	9200	-24	0,087
30	0,8917	18,66	8300	-27	0,085
31	0,8934	18,42	8400	-30	0,109
32	0,8911	18,71	8200	-27	0,097
33	0,8825	18,68	7450	-24	0,096
34	0,8929	18,75	7700	-24	0,106
35	0,8913	18,62	7500	-24	0,097
36	0,8903	18,61	7650	-24	0,106
37	0,8888	18,90	7388	-24	0,096
38	0,8917	18,38	8000	-27	0,103
39	0,8918	18,41	7400	-30	0,099
40	0,8933	19,01	7900	-27	0,110
41	0,8922	18,74	7400	-27	0,101
42	0,8919	18,98	7400	-24	0,095
43	0,8943	18,77	7300	-27	0,103
67	0,8900	20,30	-	-30	-
68	0,8900	20,64	-	-30	-
69	0,8900	20,45	-	-30	-
70	0,8900	20,76	-	-30	-
71	0,8900	20,55	-	-30	-

TAB. 5.3 Propriedades dos óleos Semi-Sintéticos e Sintéticos

Lubrificante Semi/Sintético	Densidade g/cm <sup>3</sup>	viscosidade		Ponto de Fluidez °C	Teor de Mg%(m)
		(100°C) cSt	-15°C cP		
44	0,8846	18,88	6500	-27	0,120
45	0,8823	15,32	6950	-30	0,097
46	0,8841	19,02	6200	-27	0,106
47	0,8822	15,35	6500	-30	0,102
48	0,8807	14,87	6400	-30	0,107
49	0,8857	18,47	6900	-27	0,110
50	0,8847	19,04	6600	-27	0,098
51	0,8839	18,99	6200	-27	0,108
52	0,8844	19,62	6400	-33	0,100
53	0,8823	15,08	6800	-33	0,099
54	0,8821	14,20	6600	-27	0,098
55	0,8812	14,95	6400	-33	0,102
56	0,8799	15,05	6600	-24	0,102
57	0,8493	13,70	-	-36	0,090
58	0,8514	14,15	-	-39	0,099
59	0,8513	13,97	-	-36	-
60	0,8515	14,31	-	-39	-
61	0,8506	13,88	-	-30	-
62	0,8508	13,52	-	-42	-
63	0,8523	13,74	-	-39	-
64	0,8900	20,86	-	-33	-
65	0,8900	21,02	-	-33	-
66	0,8800	20,86	-	-33	-



## 5.2 ANÁLISE DAS AMOSTRAS

Neste trabalho foram utilizadas cinco propriedades físico-químicas dos óleos lubrificantes comerciais: densidade, viscosidade a 100°C, viscosidade a -15°C, ponto de fluidez e teor de magnésio. Na TAB. 5.4 estão listadas as técnicas instrumentais e os métodos de referência empregados na determinação de tais propriedades.

TAB. 5.4 Propriedades e Métodos de Referência – Óleos Lubrificantes

<b>Propriedade</b>	<b>Método de Referência</b>	<b>Técnica Instrumental</b>
Densidade	ASTM 4052	Densímetro Digital
Viscosidade Brookfield (-15 °C)	ASTM D2893	Viscosímetro Brookfield
Viscosidade (100 °C)	ASTM D445	Viscosímetro Cinemático
Ponto de Fluidez	ASTM D97	Método Normalizado
Teor de Magnésio	ASTM D4628	Absorção Atômica

## 5.3 CONJUNTO DE AMOSTRAS

Para se desenvolver um método analítico por espectroscopia no infravermelho é necessário que as amostras disponíveis sejam separadas em dois grupos: um grupo para o conjunto de modelagem ou calibração e um grupo para validação ou previsão.

O grupo de calibração é utilizado para a construção ou geração do modelo e o de previsão é utilizado para a validação (externa) de tal modelo.

Neste trabalho, a habilidade de previsão dos modelos de calibração no infravermelho médio foi estimada usando-se o método de ordenamento para seleção de amostras para as seguintes propriedades: densidade, viscosidade a 100°C e ponto de fluidez.

No caso da espectroscopia de infravermelho próximo (NIR) e para as propriedades do MIR: teor de magnésio e viscosidade a -15°C foi usado o método da validação cruzada devido ao pequeno número de amostras para a construção dos modelos de calibração.

## 5.4 LEITURA DOS ESPECTROS

### 5.4.1 ESPECTROS NIR

Os espectros de infravermelho próximo (NIR) foram obtidos em um espectrofotômetro Varian Cary 500, cedido pelo Laboratório de Catálise do Instituto Nacional de Tecnologia (LACAT-INT). A sala onde se encontra o espectrofotômetro tem a temperatura controlada em torno de 25°C.

As análises das amostras foram realizadas sem nenhum preparo prévio, em cubetas de plástico descartável com 1,0 cm de caminho ótico.

Os espectros foram adquiridos na região de 800 a 2240 nm em escala de absorbância. Esta região foi escolhida baseada no trabalho de LIMA et al (2005) que a empregou para análise de óleos lubrificantes básicos.

Com o propósito de eliminar ruídos de diferentes fontes, foi realizada uma correção de fundo (*background*), que consiste em adquirir e arquivar um espectro de um branco, do ar neste caso, para posteriormente subtraí-lo dos espectros de amostras a serem adquiridos.

Após o término da leitura, os espectros foram salvos no formato ASCII para facilitar a importação dos dados para uma planilha eletrônica, e, posteriormente, para o programa comercial UNSCRABLER versão 6.11 CAMO; este programa é um pacote estatístico especificamente desenvolvido para quimiometria e, em particular, para o processamento de dados espectrais.

### 5.4.2 ESPECTROS MIR

Os espectros de infravermelho médio (MIR) foram adquiridos em um aparelho FT-IR Perkin-Elmer 1710 com detetor DTGS (Deuterated TriGlycine Sulfate). As amostras foram colocadas em uma célula selada de KBr de espessura fixa de 0.025 mm (FIG 5.1). A escolha da janela de KBr deve-se à sua transparência na região utilizada para a leitura dos espectros e seu baixo custo.



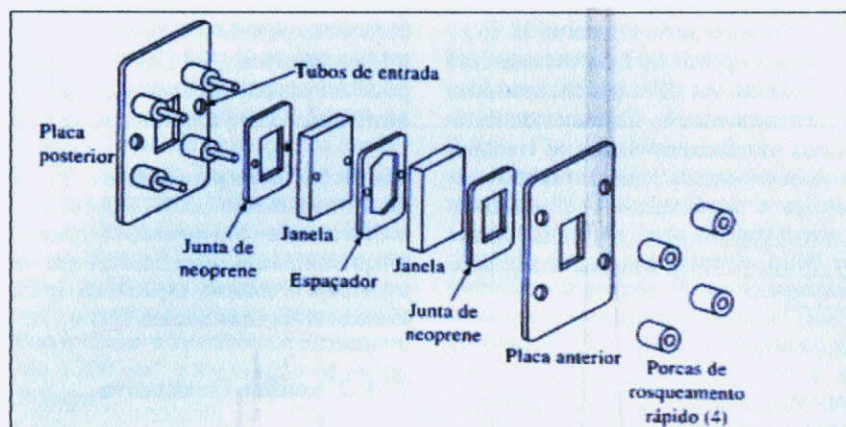


FIG. 5.1 Modelo de célula fixa de KBr usado no MIR

Um dos problemas relacionados com a espectroscopia MIR dos lubrificantes é a alta absorbância em algumas regiões do espectro. Devido a limitações da Lei de Beer, só foram utilizados dados espectrais com valores de absorbância inferiores a duas unidades. Isto porque em altas concentrações, a distancia média entre as moléculas responsáveis pela absorção diminui a ponto de cada molécula afetar a distribuição de carga da sua vizinha, o que pode alterar a capacidade das moléculas de absorver um determinado comprimento (SKOOG, 2002). A FIG. 5.2 apresenta um espectro de infravermelho de um óleo lubrificante analisado e mostra as altas absorbâncias da região espectral em torno de  $2800-3200\text{ cm}^{-1}$  onde a não linearidade é um problema.

Para este trabalho, os espectros foram lidos na região de  $400-2000\text{ cm}^{-1}$  (circulo da FIG.5.2).

Foi realizada a correção de *background*, considerando-se o espectro do ar como branco.

Cada espectro foi lido com 32 varreduras aditivas com resolução  $2\text{ cm}^{-1}$  em escala de absorbância. O tempo de leitura nesse caso é aproximadamente 7 minutos. Levando em consideração o tempo de preparo, o procedimento para cada amostra não ultrapassa 15 minutos.

Os espectros foram salvos em formato ASCII que contém a leitura individual do espectro de cada amostra, além de informações extras como instrumento usado, nome da amostra, etc. Criando uma rotina para a importação dos dados para o UNSCRABLER versão 6.11 Camo, utilizado para as análises de regressão, análises de componentes principais e pré-tratamento.

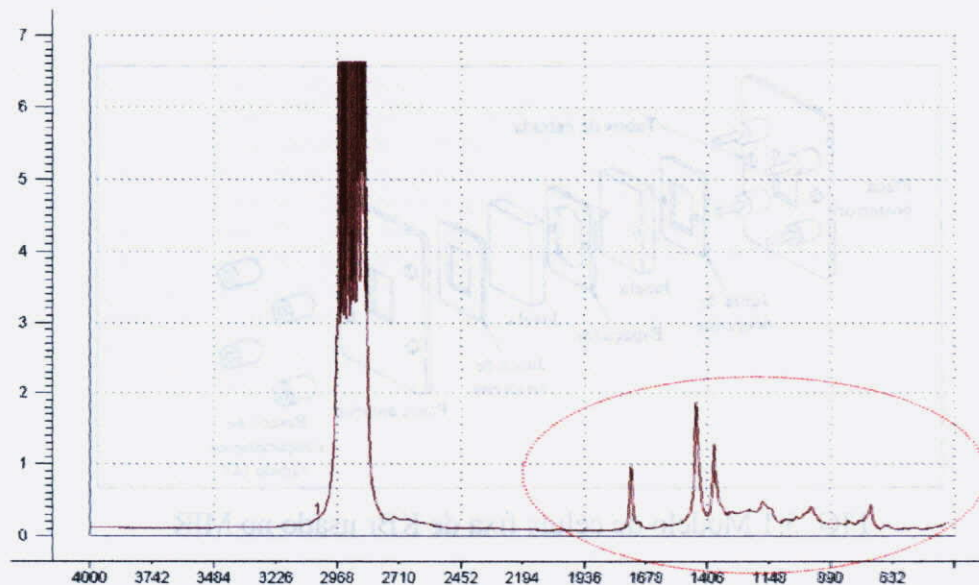


FIG. 5.2 Espectro de óleo lubrificante no MIR com caminho ótico 0,025 mm.

## 5.5 TRATAMENTO DOS DADOS

### 5.5.1 DESENVOLVIMENTO DOS MODELOS

O método de análise de regressão empregado para o desenvolvimento dos modelos foi o dos mínimos quadrados parciais (PLS). O método escolhido para a otimização dos modelos assim obtidos foi o de validação cruzada completa (*leaving-one-out full cross validation*) de acordo com o item 4.3.2.

A identificação e subsequente eliminação das amostras anômalas ou aberrantes na análise de regressão foram realizadas mediante a inspeção visual do gráfico de influência tridimensional. Neste gráfico se encontram lançados os valores de *leverage* versus os quadrados dos resíduos, tanto para a variável dependente quanto para a variável independente. As amostras que apresentaram simultaneamente altos valores de *leverage* e resíduo foram eliminadas do conjunto de calibração.

A decisão do número de variáveis latentes (ou fatores) a serem empregadas em cada modelo de calibração foi realizada através da inspeção visual do gráfico do RMSEC versus número de variáveis latentes. Para tanto, realizam-se várias regressões, uma para cada número



de variáveis latentes, e calcula-se o respectivo valor de RMSEC. Obtém-se, por conseguinte, um modelo e um valor de RMSEC para cada número de variáveis latentes. Em seguida, constrói-se o gráfico do RMSEC versus número de variáveis latentes; o número ótimo de variáveis latentes é aquele depois do qual o valor de RMSEC não mais decresce de modo significativo.

## 5.5.2 SELEÇÃO DAS VARIÁVEIS INDEPENDENTES

Neste trabalho, foram empregados dois métodos de seleção de variáveis independentes, para avaliar quais os comprimentos de onda de um espectro que efetivamente devem ser usados na análise de regressão para o desenvolvimento dos modelos de calibração: análise de componentes principais (PCA) e coeficiente de correlação.

No item 5.3.1.2 foi descrita a análise de componentes principais. O segundo método foi aplicado da seguinte forma: as correlações entre todas as variáveis são armazenadas em uma matriz de correlação, onde obtemos um valor do coeficiente de correlação entre as propriedades (variáveis dependentes) e todos os comprimentos de onda (variáveis independentes); de acordo com os valores de coeficiente de correlação, selecionamos os comprimentos de onda com maior correlação, que serão empregados nos modelos de calibração (OTTO,1999).

Para calcularmos a matriz de correlação usamos o **coeficiente de correlação**, de acordo com as equações abaixo:

$$r_{jk} = \frac{\text{cov}(j,k)}{s_j s_k} \quad j \neq k \quad (\text{EQ. 6.1})$$

$$s_j = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}} \quad (\text{EQ. 6.2})$$

$$s_k = \sqrt{\frac{\sum_{i=1}^n (x_{ik} - \bar{x}_k)^2}{n-1}} \quad (\text{EQ. 6.3})$$

$$\text{cov}(j,k) = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \quad (\text{EQ. 6.4})$$

Onde:

$S_j$  e  $S_k$  = são os desvios padrões;

$\text{cov}(j,k)$  = covariância para  $j, k = 1 \dots p; \quad j \neq k$

Dessa forma, a matriz de correlação pode ser feita como mostrado abaixo:

$$R = \begin{bmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{12} & 1 & & r_{2p} \\ \vdots & & & \vdots \\ r_{1p} & r_{2p} & \dots & 1 \end{bmatrix}$$

Usamos o termo **correlação positiva** quando  $r > 0$  (nesse caso à medida que **X** cresce também cresce **Y**), e **correlação negativa** quando  $r < 0$  (nesse caso à medida que **X** cresce, **Y** decresce).

Quanto maior o valor do módulo de  $r$ , mais forte a associação. No extremo, se  $r = 1$  ou  $r = -1$ , todos os pontos no gráfico de dispersão caem exatamente numa linha reta. No outro extremo, se  $r = 0$ , não existe nenhuma associação linear.

O seguinte quadro fornece um guia de como podemos descrever a correlação em palavras, dado o valor numérico. É claro que as interpretações dependem de cada contexto em particular.



TAB. 5.5 Interpretação dos valores de  $r^2$ .

Valor de r (+ ou -)	Interpretação
0,00 a 0,19	Uma correlação bem fraca
0,20 a 0,39	Uma correlação fraca
0,40 a 0,69	Uma correlação moderada
0,70 a 0,89	Uma correlação forte
0,90 a 1,00	Uma correlação muito forte

## 6.0 RESULTADOS E DISCUSSÃO

### 6.1 SELEÇÃO DAS AMOSTRAS E PROPRIEDADES

A primeira etapa deste trabalho se consistiu na aquisição das amostras de óleos lubrificantes comerciais em postos de gasolina e em outros estabelecimentos comerciais. Os valores primários das propriedades foram gentilmente cedidos pelos fabricantes BR distribuidora e PROMAX BARDAHL.

As propriedades para as quais se desenvolveram os modelos de calibração foram: densidade, viscosidade a 100 °C, viscosidade a -15 °C, ponto de fluidez e teor de magnésio. Algumas estatísticas destes dados são apresentadas na TAB. 6.1.

TAB. 6.1 Estatística dos dados – óleos Lubrificantes Comerciais

Propriedades	Amplitude	Desvio Padrão	Mediana	Média
Densidade	0,0470	0,0123	0,8900	0,8857
Viscosidade a 100°C	7,498	2,545	18,415	17,098
Viscosidade a -15°C	4250	951	7500	7536
Teor de Magnésio	0,047	0,008	0,099	0,098
Ponto de Fluidez	33	7	-27	-26

### 6.2 ESPECTROS DAS AMOSTRAS

Na faixa selecionada para o espectro de infravermelho próximo, entre 2240 – 800 nm (FIG. 6.1), é possível observar a presença de uma banda de média intensa ao redor de 1210 nm (referentes às vibrações axiais de segundo sobretom de ligações C-H) e de duas bandas de média intensidade entre 1390 e 1420 nm (referente à combinação do primeiro sobretom da vibração axial com a vibração angular fundamental de ligações C-H alifático). Além disso, formam-se duas bandas entre 1690 e 1780 nm com grande intensidade, devido à sobreposição das bandas geradas pelas vibrações de primeiro sobretom de ligações C-H parafínicas, olefínicas e aromáticas e pelas vibrações de deformação axial C=O dos ésteres. Observa-se



também uma banda em torno de 2180 nm, que pode ser atribuída à combinação da vibração axial fundamental com a vibração angular fundamental de ligações C-H aromáticas.

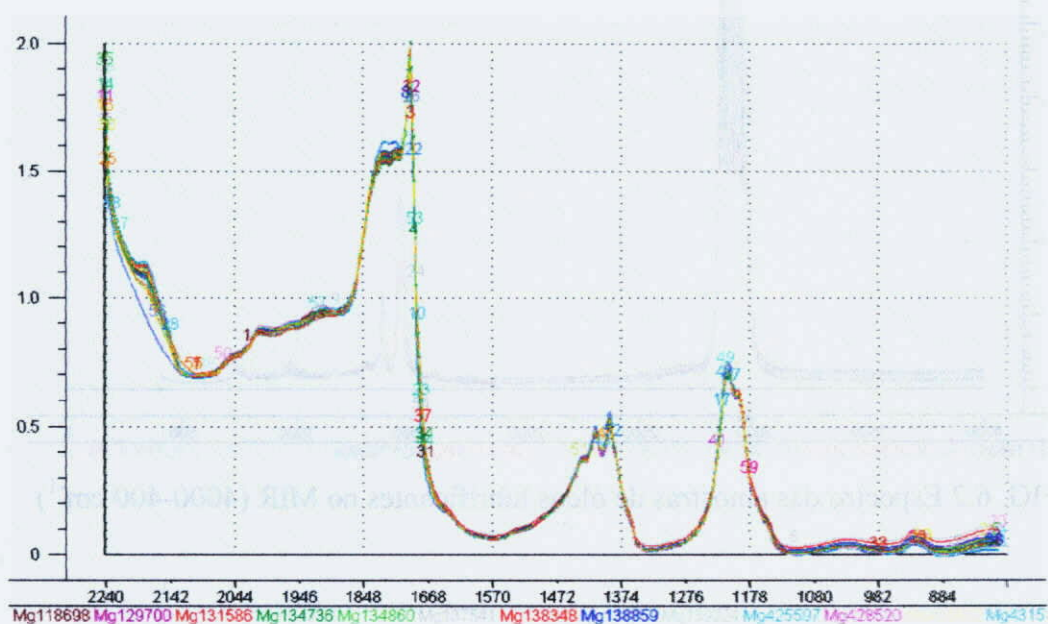


FIG. 6.1 Espectro das amostras de óleos lubrificantes no NIR (2240-800 nm)

Já na faixa selecionada para o espectro de infravermelho médio (FIG. 6.2), pode-se observar a presença das bandas de hidrocarbonetos saturados referentes às vibrações fundamentais de estiramento e deformação de ligações C-H alifáticas, entre  $3000 - 2850 \text{ cm}^{-1}$  e entre  $1463-1377 \text{ cm}^{-1}$ , respectivamente. Pode-se observar também uma banda de média intensidade em torno de  $721 \text{ cm}^{-1}$ , devido às vibrações fundamentais do tipo rock de ligações C-H, onde o número de átomos de carbono em cadeia é maior ou igual a quatro, e devido às vibrações de deformação axial C=O (entre  $1776-1735 \text{ cm}^{-1}$ ) e C-O (entre  $1300-1000 \text{ cm}^{-1}$ ) dos ésteres.

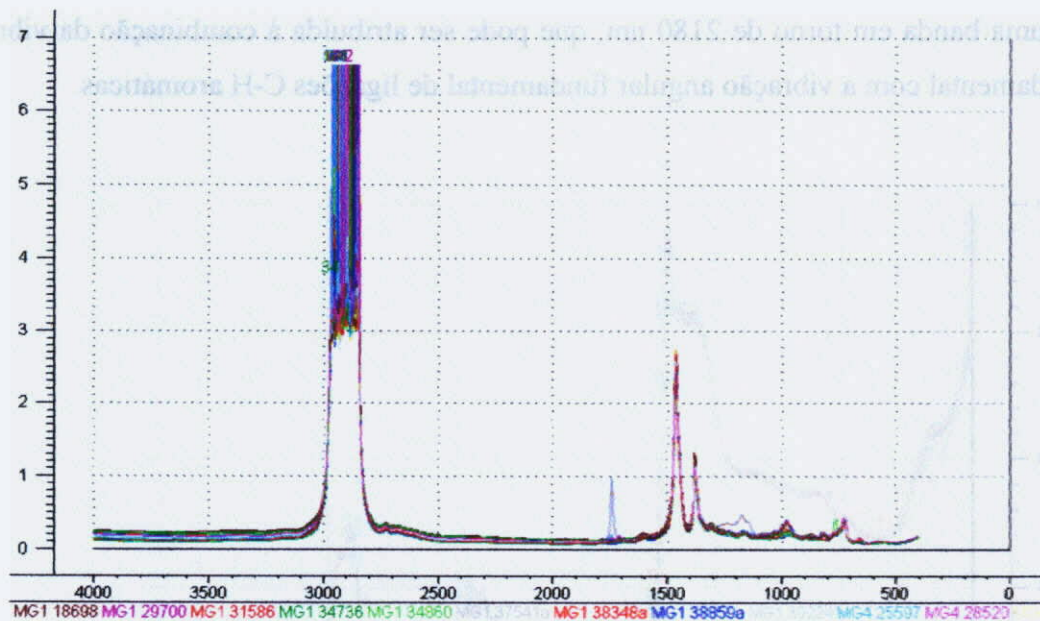


FIG. 6.2 Espectro das amostras de óleos lubrificantes no MIR (4000-400  $\text{cm}^{-1}$ )

### 6.3 CLASSIFICAÇÃO DAS AMOSTRAS

Na análise da PCA, os *scores* são as coordenadas das amostras em um novo sistema de coordenadas, cujos eixos são definidos pelos PCs. Os novos eixos auxiliam na visualização dos dados originais em um número menor de dimensões. Assim, os gráficos de *scores* revelam a relação entre as amostras, de forma que quanto mais próximas as amostra estiverem mais semelhantes serão.

#### 6.3.1 APLICAÇÃO DA PCA NOS DADOS DAS PROPRIEDADES

Primeiramente, a análise de componentes principais (PCA) foi realizada com a finalidade de observar o comportamento das propriedades físico-químicas determinadas para cada uma das 71 amostras dos óleos lubrificantes comerciais. As propriedades físico-químicas analisadas foram: densidade, viscosidade a 100 °C e Ponto de Fluidez.

As amostras foram numeradas de acordo com os tipos de óleos e indústrias.



Minerais: tipo A<sub>1</sub>, de 1-11; tipo A<sub>2</sub>, de 12-19; tipo A<sub>3</sub>, de 20-32; tipo A<sub>4</sub>, de 33-43 e tipo B<sub>1</sub>, de 67-71;

Semi-sintéticos: Tipo A<sub>5</sub>, de 44 a 56;

Sintéticos: Tipo A<sub>6</sub>, de 57 a 63; e Tipo B<sub>2</sub>, de 64 a 66.

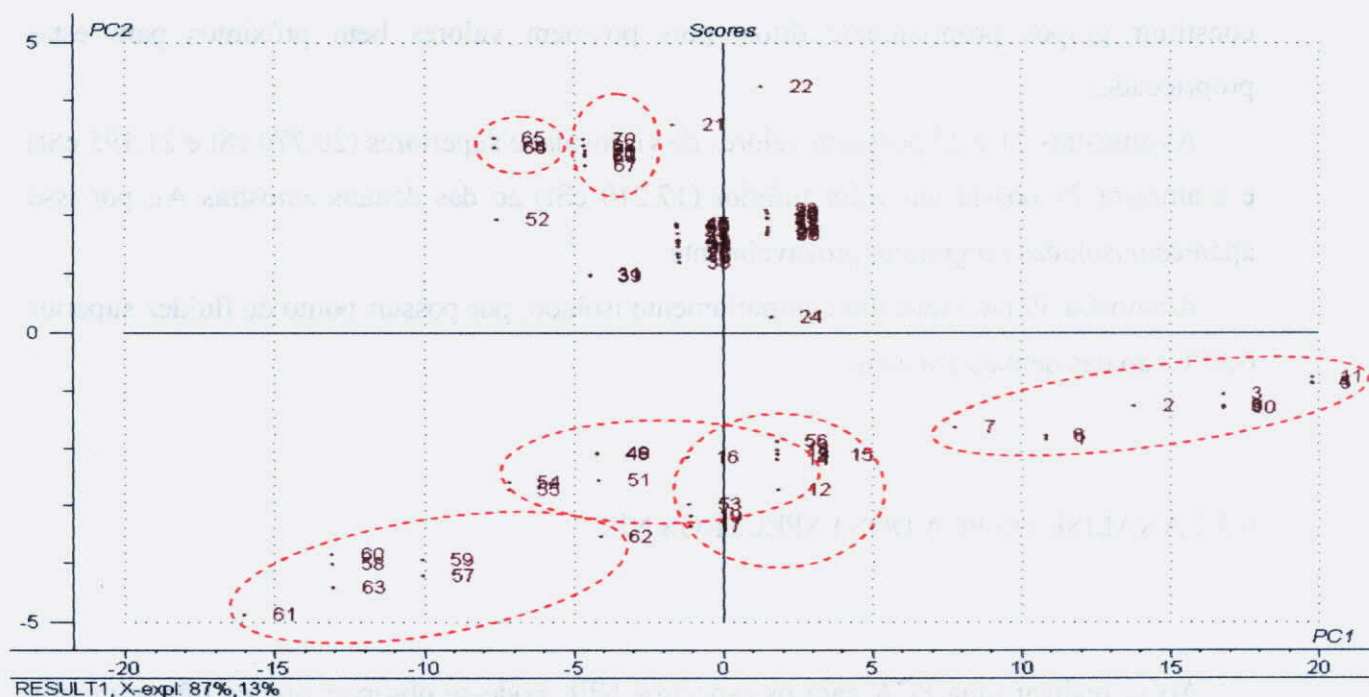


FIG. 6.3 Gráfico de scores PC<sub>1</sub> X PC<sub>2</sub> com todas as 71 amostras.

Apenas dois componentes principais são capazes de explicar 100% (87% PC<sub>1</sub>, mas 13% PC<sub>2</sub>) da variância entre as amostras de óleos lubrificantes estudadas.

O gráfico de *scores* entre PC<sub>1</sub> e PC<sub>2</sub> (FIG. 6.3) mostra que as amostras foram quase todas separadas em seus grupos originais.

As propriedades das amostras estudadas se encontram nas TAB.s 5.2 e 5.3.

As amostras dos tipos A<sub>2</sub>, B<sub>1</sub>, A<sub>6</sub> e B<sub>2</sub> formaram grupos totalmente isolados uns dos outros, por possuírem propriedades físico-químicas bem distintas entre si. Somente a amostra 62 ficou afastada do grupamento A<sub>6</sub>, por possuir um ponto de fluidez inferior (-30 °C) aos das demais amostras sintéticas.

As amostras do tipo A<sub>1</sub> possuem pontos de fluidez inferiores aos das demais amostras. Além disso, esses valores são bastante dispersos entre si, constituindo um grupo bastante espalhado, que pode ser observado no lado direito do gráfico.

Apenas parte das amostras semi-sintéticas formou um grupamento. Além disso, duas amostras (53 e 56) ficaram juntas do grupamento das amostras A<sub>2</sub>, por possuírem propriedades semelhantes a essas, e a amostra 52 apresentou-se isolada do grupamento, por possuir uma viscosidade superior (19,62 cSt) a das demais amostras A<sub>5</sub>.

As demais amostras A<sub>5</sub> se agruparam com as amostras A<sub>3</sub> e A<sub>4</sub> de forma aleatória, sem constituir grupos propriamente ditos, pois possuem valores bem próximos para estas propriedades.

As amostras 21 e 22 possuem valores de viscosidade superiores (20,770 cSt e 21,195 cSt) e a amostra 24 possui um valor inferior (17,210 cSt) ao das demais amostras A<sub>3</sub>, por isso aparecem isoladas no gráfico, provavelmente.

A amostra 39 (A<sub>4</sub>) teve um comportamento isolado, por possuir ponto de fluidez superior (-30°C) ao das demais amostras.

### 6.3.2 ANÁLISE DE PCA DOS ESPECTROS NIR

Ao se realizar uma PCA para os espectros NIR, pode-se observar que os três primeiros PCs explicam 95% da variância original contida nas 680 variáveis originais, enquanto que os seis primeiros respondem por 99% da variação contida nos dados originais. As FIG.s 6.4 a 6.7 mostram os gráficos de *scores* com maior classificação de amostras para a matriz espectral original e para as matrizes autoescaladas, normalizadas e derivada após normalização, respectivamente.

Esta análise foi realizada com os espectros NIR das 60 amostras de óleos lubrificantes empregadas na construção dos Modelos.

As amostras foram separadas por tipos de óleos e indústrias:

Mínerais: tipo A<sub>1</sub>, de 1-9; tipo A<sub>2</sub>, de 10-17; tipo A<sub>3</sub>, de 18-26; tipo A<sub>4</sub>, de 27-36; e tipo B<sub>1</sub>, de 56-60;

Semi-sintéticos: Tipo A<sub>5</sub>, de 37-46;

Sintéticos: Tipo A<sub>6</sub>, de 47-52; e Tipo B<sub>2</sub>, de 53-55.

Ao compararmos esses gráficos com o da FIG. 6.3, podemos observar que, em termos de conteúdo, ambos separam as amostras em seus respectivos grupamentos. Entretanto, aqui analisamos a composição química (estrutura) das amostras.



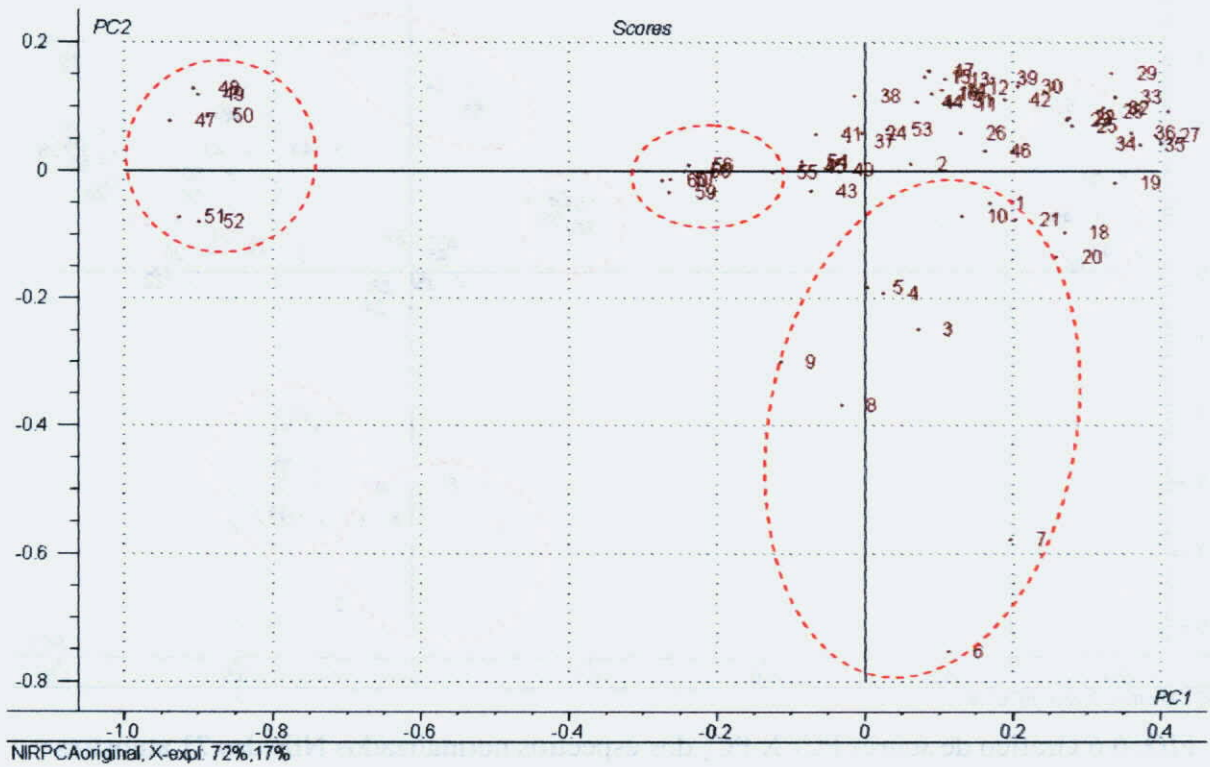


FIG. 6.4 Gráfico de *scores* PC<sub>1</sub> X PC<sub>2</sub> dos espectros originais NIR das 71 amostras

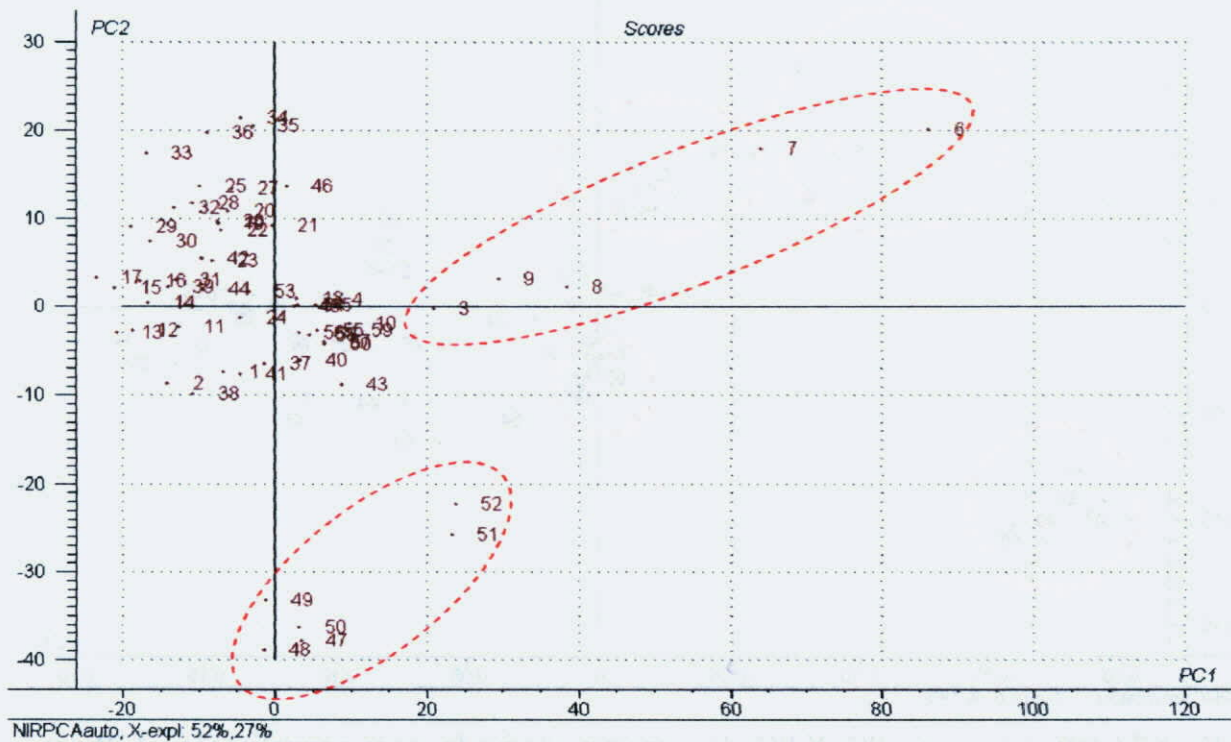


FIG. 6.5 Gráfico de *scores* PC<sub>1</sub> X PC<sub>2</sub> dos espectros autoescalados NIR das 71 amostras

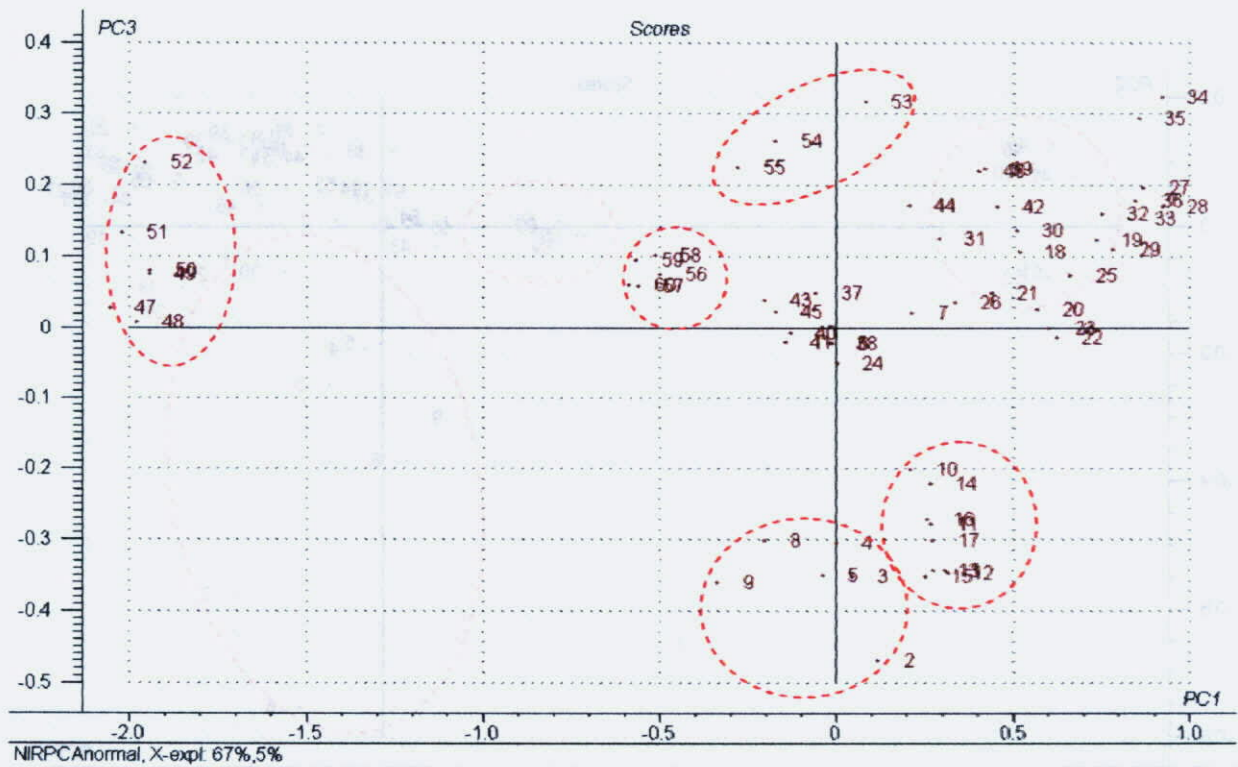


FIG. 6.6 Gráfico de *scores* PC<sub>1</sub> X PC<sub>2</sub> dos espectros normalizados NIR das 71 amostras

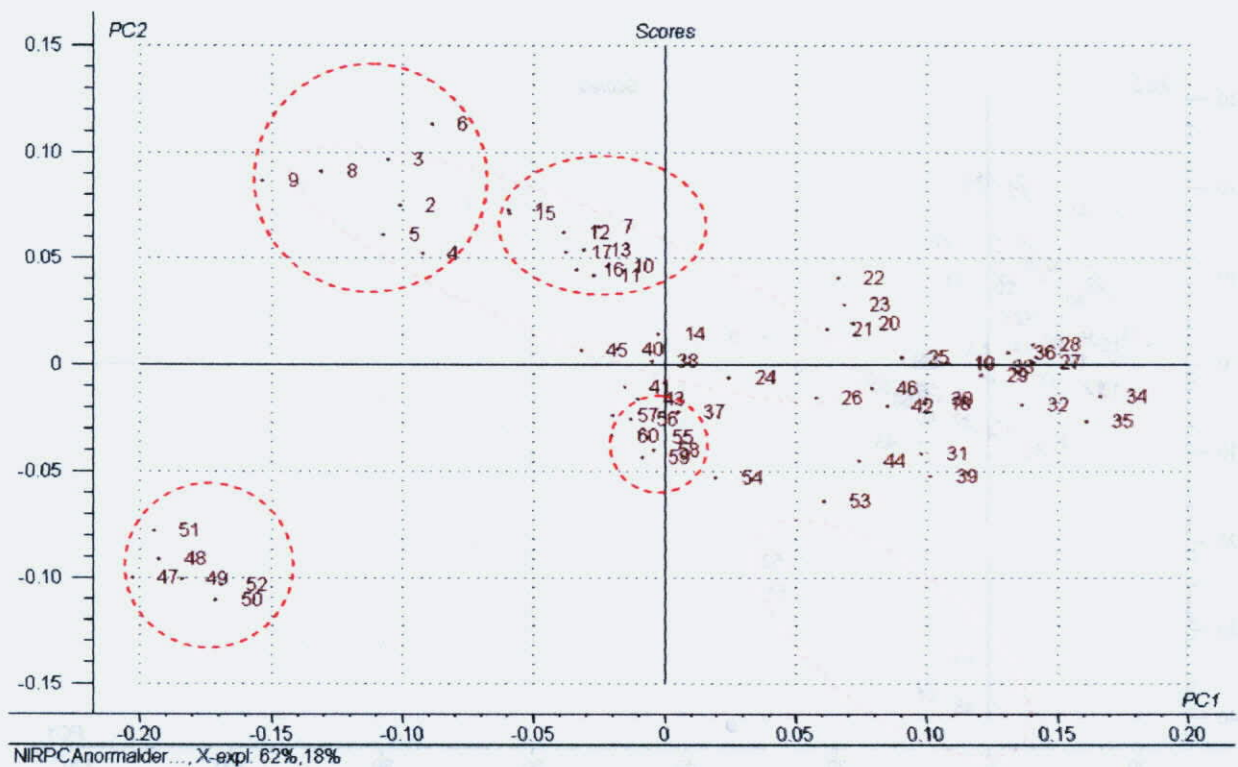


FIG. 6.7 Gráfico de *scores* PC<sub>1</sub> X PC<sub>2</sub> dos espectros derivados após normalização e NIR das 71 amostras



Quando se analisa os gráficos obtidos, pode-se verificar que a normalização dos espectros foi o pré-tratamento que melhor separou as amostras estudadas, distinguindo todas as amostras A<sub>2</sub>, B<sub>1</sub>, A<sub>6</sub>, B<sub>2</sub> e 78% das amostras A<sub>1</sub>. As demais amostras se distribuem aleatoriamente, sem constituir grupos definidos.

### 6.3.3 ANÁLISE DE PCA DOS ESPECTROS MIR

Ao se realizar uma PCA para aos espectros no MIR, pode-se observar que os cinco primeiros PCs explicam 98% da variância original, e que os seis primeiros respondem por 99% da variância contida nos dados originais. Da mesma forma que os gráficos de *scores* no NIR, as FIG. 6.8 a 6.11 mostram os gráficos de *scores* com maior classificação de amostras para a matriz espectral original e para a matriz autoescalorada, normalizada e derivada após normalização, respectivamente.

Esta análise foi realizada com os espectros MIR das 71 amostras de óleos lubrificantes empregadas na construção dos modelos.

Nestes gráficos, as amostras também foram separadas por tipos de óleos:

- Minerais: Tipo A<sub>1</sub>, de 1-11; Tipo A<sub>2</sub>, de 12-19; Tipo A<sub>3</sub>, de 20-32; Tipo A<sub>4</sub>, de 33-43; e Tipo B<sub>1</sub>, de 67-71;
- Semi-sintéticos: Tipo A<sub>5</sub>, de 44-56;
- Sintéticos: Tipo A<sub>6</sub>, de 57-63; e Tipo B<sub>2</sub>, de 64-66.

Comparando-se os gráficos da FIG. 6.8 até a FIG. 6.11 com o da FIG. 6.3, nota-se que ambos separam as amostras em seus respectivos grupamentos, entretanto aqui o enfoque está voltado para a composição química das amostras.

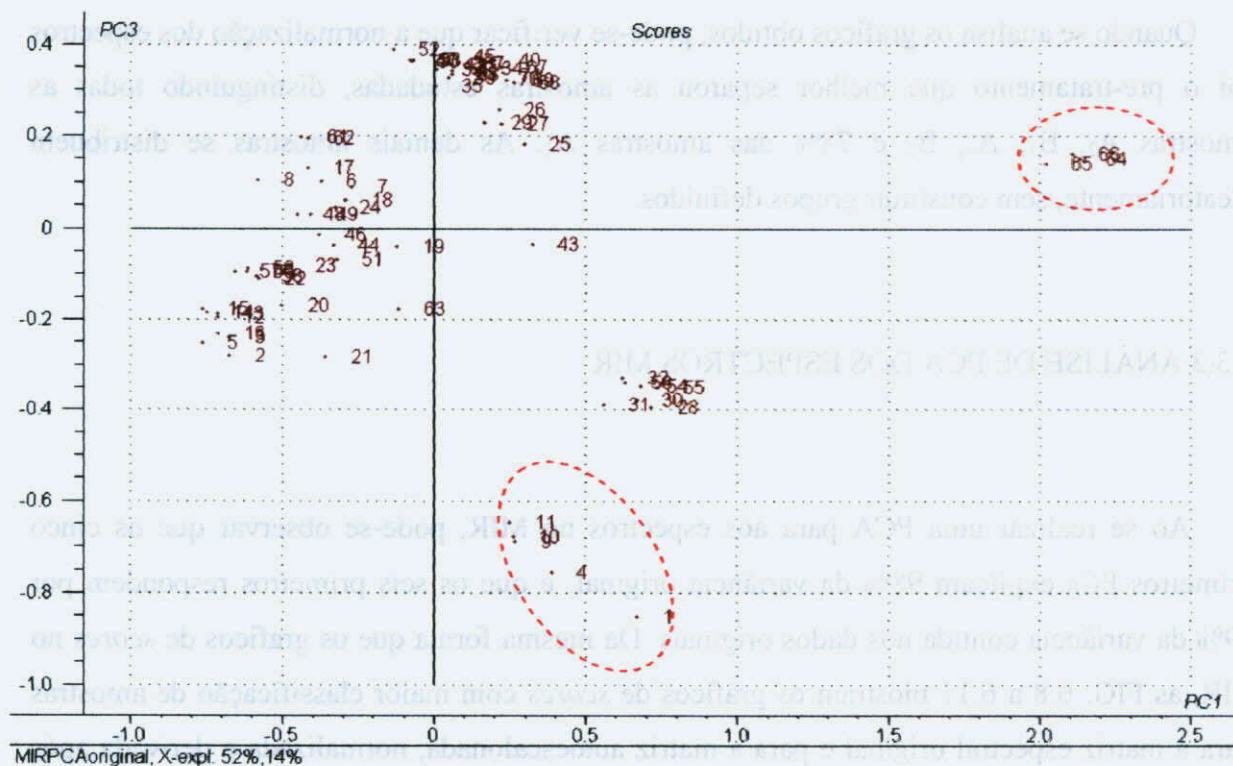


FIG. 6.8 Gráfico de *scores* PC<sub>1</sub> X PC<sub>2</sub> dos espectros originais MIR das 71 amostras

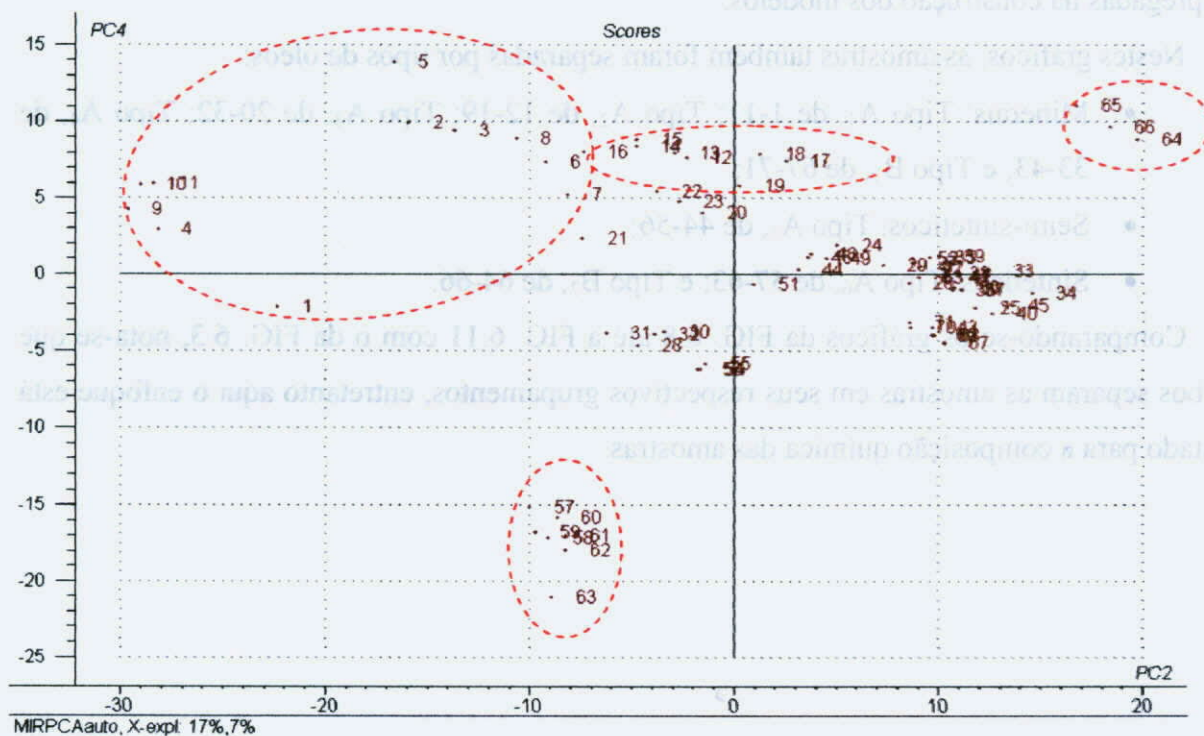


FIG. 6.9 Gráfico de *scores* PC<sub>1</sub> X PC<sub>2</sub> dos espectros autoescalados MIR das 71 amostras



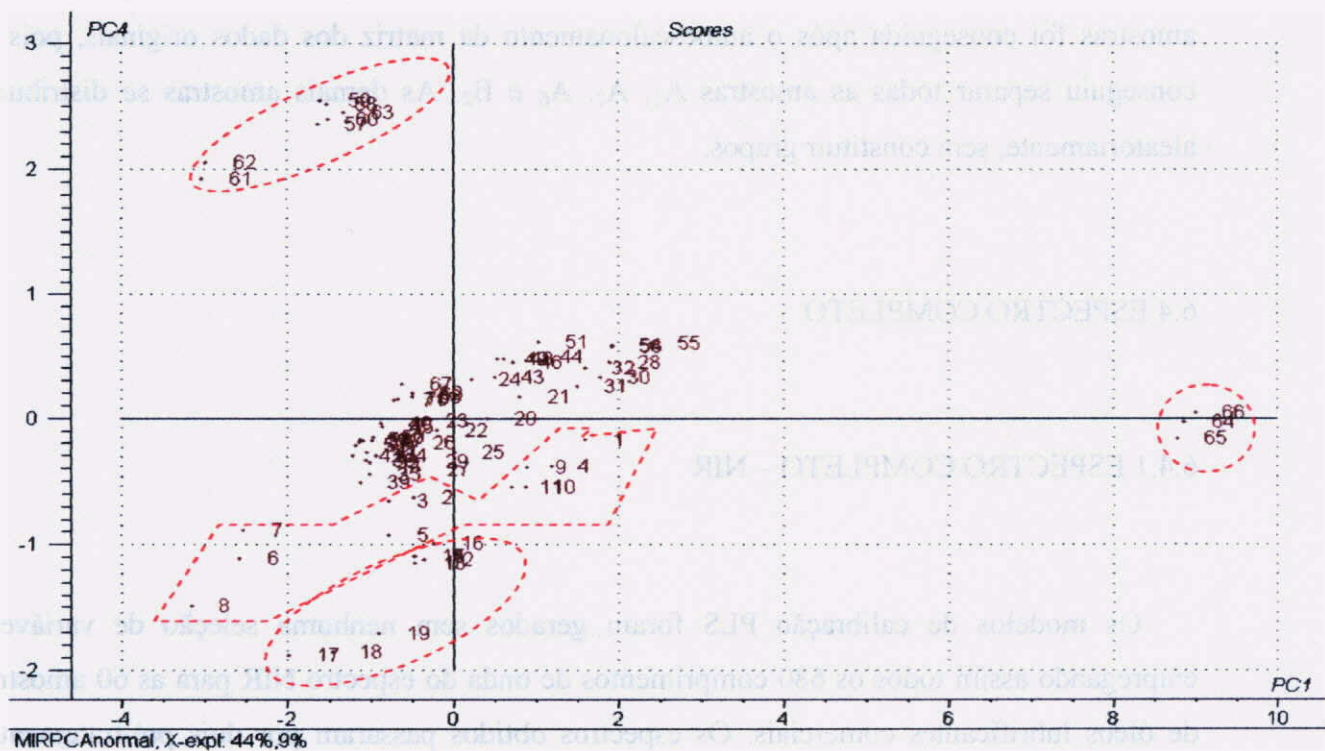


FIG. 6.10 Gráfico de *scores* PC<sub>1</sub> X PC<sub>2</sub> dos espectros normalizados MIR das 71 amostras

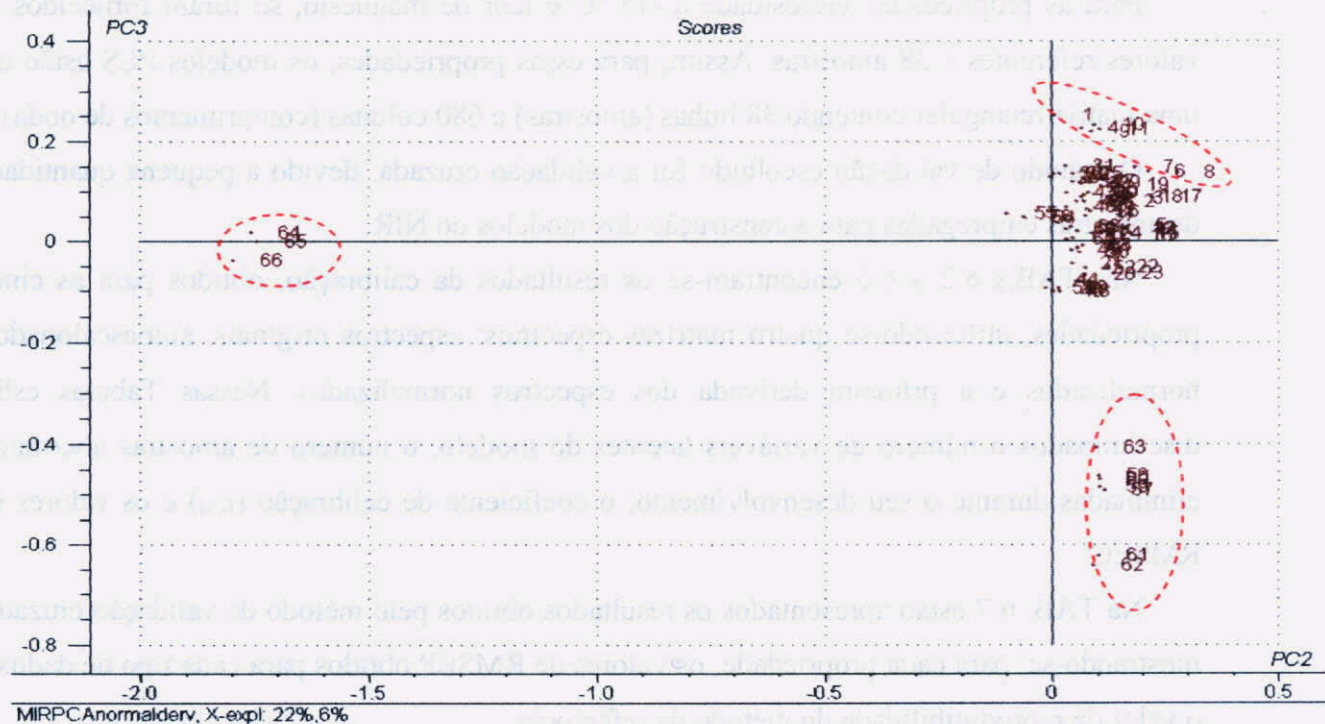


FIG. 6.11 Gráfico de *scores* PC<sub>1</sub> X PC<sub>2</sub> dos espectros derivados após normalização MIR das 71 amostras

Ao analisarmos os gráficos obtidos, podemos verificar que a melhor separação das amostras foi conseguida após o autoescalonamento da matriz dos dados originais, pois se conseguiu separar todas as amostras A<sub>1</sub>, A<sub>2</sub>, A<sub>6</sub> e B<sub>2</sub>. As demais amostras se distribuem aleatoriamente, sem constituir grupos.

## 6.4 ESPECTRO COMPLETO

### 6.4.1 ESPECTRO COMPLETO – NIR

Os modelos de calibração PLS foram gerados sem nenhuma seleção de variáveis, empregando assim todos os 680 comprimentos de onda do espectro NIR para as 60 amostras de óleos lubrificantes comerciais. Os espectros obtidos passaram por dois pré-tratamentos iniciais: os valores de transmitância foram transformados em absorbância e em seguida centrados na média.

Para as propriedades viscosidade a -15 °C e teor de magnésio, só foram fornecidos os valores referentes a 38 amostras. Assim, para essas propriedades, os modelos PLS estão em uma matriz retangular contendo 38 linhas (amostras) e 680 colunas (comprimentos de onda).

O método de validação escolhido foi à validação cruzada, devido à pequena quantidade de amostras empregadas para a construção dos modelos no NIR.

Nas TAB.s 6.2 a 6.6 encontram-se os resultados da calibração, obtidos para as cinco propriedades, utilizando-se quatro matrizes espectrais: espectros originais, autoescalonados, normalizados e a primeira derivada dos espectros normalizados. Nessas Tabelas estão discriminados o número de variáveis latentes do modelo, o número de amostras aberrantes eliminadas durante o seu desenvolvimento, o coeficiente de calibração ( $r_{cal}$ ) e os valores de RMSEC.

Na TAB. 6.7 estão apresentados os resultados obtidos pelo método de validação cruzada, mostrando-se, para cada propriedade, os valores de RMSEP obtidos para cada tipo de dados e o valor da reprodutibilidade do método de referência.

A incerteza de um método de análise instrumental secundário é calculada como  $\pm 2$  vezes o valor do RMSEP deste método. Além disso, um método secundário não pode ter uma



precisão maior que a do método primário de referência. Assim, um método de análise instrumental secundário, como a espectroscopia no NIR, só se torna útil para substituir o método de referência primário se o RMSEP obtido para as amostras de validação for, no máximo, igual à metade da reprodutibilidade do método de referência. (LIMA, 2003).

Entretanto, se os modelos apresentarem um erro relativo baixo, podem ser empregados no controle de processos, onde são admitidas maiores margens de erro (BORIN,2003).

Para o método ASTM D 445 da viscosidade a 100 °C, o valor da reprodutibilidade é 0,76 % do valor medido, enquanto que para a viscosidade a -15 °C, método D 5293, o valor da reprodutibilidade é 8,9 % do valor medido. Assim, para essas propriedades, fez-se necessário calcular o valor da reprodutibilidade utilizando o valor médio (TAB. 6.1) como valor medido.

TAB. 6.2 PLS no NIR – Densidade. Espectro Completo. Validação: Validação Cruzada.

Densidade	Variáveis Latentes	Aberrantes	r <sub>cal</sub>	RMSEC
Original	5	14,27,38	0,9869	0,0020
Autoescalados	5	27,55	0,9925	0,0015
Normalizado	6	27,55	0,9935	0,0014
Normalizado e Derivado	5	27,55	0,9938	0,0014

TAB. 6.3 PLS no NIR- Viscosidade a 100°C. Espectro Completo. Validação: Validação Cruzada.

Viscosidade a 100 °C	Variáveis Latentes	Aberrantes	r <sub>cal</sub>	RMSEC
Original	12	9,20,46	0,9865	0,42
Autoescalados	11	31,44,45	0,9852	0,45
Normalizado	11	9,23,44,46	0,9840	0,46
Normalizado e Derivado	10	9,10,20,36	0,9872	0,41

TAB. 6.4 PLS no NIR- Viscosidade a -15 °C. Espectro Completo. Validação: Validação Cruzada.

Viscosidade a -15 °C	Variáveis Latentes	Aberrantes	r <sub>cal</sub>	RMSEC
Original	9	4,7,12	0,9786	176
Autoescalados	8	4,7,12	0,9748	191
Normalizado	12	4,7,12,23	0,9899	123
Normalizado e Derivado	7	4,7,10,27	0,9775	178



TAB. 6.5 PLS no NIR- Ponto de Fluidez. Espectro Completo. Validação: Validação Cruzada.

Ponto de Fluidez	Variáveis Latentes	Aberrantes	$r_{cal}$	RMSEC
Original	10	21,46,49,52	0,9790	1,56
Autoescalados	6	15,33,45,52	0,9602	2,15
Normalizado	9	17,33,45,52	0,9747	1,72
Normalizado e Derivado	6	15,31,45,52	0,9647	2,04

TAB. 6.6 PLS no NIR- Teor de Magnésio. Espectro Completo. Validação: Validação Cruzada.

Teor de Magnésio	Variáveis Latentes	Aberrantes	$r_{cal}$	RMSEC
Original	6	12,21,25,28	0,8516	0,003
Autoescalados	7	21,25,28,30	0,8695	0,004
Normalizado	6	12,21,25,28	0,8508	0,003
Normalizado e Derivado	6	10,27,31,35	0,9053	0,004

TAB. 6.7 Resultados da Validação no NIR. Espectros Completos. Validação: Validação Cruzada

Propriedades	RMSEP				Reproduz.
	Esp. Orig.	Esp. auto.	Esp. Norm.	Esp. N1D.	
Densidade	0,0025 g/cm <sup>3</sup>	0,0018 g/cm <sup>3</sup>	0,0017 g/cm <sup>3</sup>	0,0018 g/cm <sup>3</sup>	0,0005g/cm <sup>3</sup>
Viscosidade a 100 °C	0,65 cSt	0,69 cSt	0,71 cSt	0,63 cSt	0,13 cSt
Viscosidade a -15 °C	344 cP	312 cP	298 cP	313 cP	670 cP
Ponto de Fluidez	2 °C	3 °C	3 °C	3 °C	6 °C
Teor de Magnésio	0,005 %(m)	0,006 %(m)	0,005 %(m)	0,006 %(m)	0,032 %(m)

#### 6.4.1.1 DISCUSSÃO - ESPECTRO COMPLETO (NIR)

Para a densidade, o melhor modelo foi observado com os dados normalizados. Entretanto, os valores de RMSEP para os dados autoescalados e derivados após normalização (primeira derivada dos espectros normalizados) foram muito próximos do valor encontrado para a normalização.

Para a viscosidade a -15 °C, o menor valor de RMSEP foi obtido utilizando a normalização como pré-tratamento dos dados originais.

Já para a viscosidade a 100 °C, o menor resultado de RMSEP foi obtido com a primeira derivada dos espectros normalizados.

Para o teor de magnésio, foram encontrados valores iguais de RMSEP tanto para os dados originais quanto para os dados normalizados.



Já para o ponto de fluidez, obteve-se o menor valor de RMSEP com os dados originais, demonstrando que, para essa propriedade, o impacto das variações aleatórias e não uniformes não interferem de modo significativo no resultado.

Comparando-se os valores de RMSEP com os da reprodutibilidade, observa-se que, nos casos da densidade, viscosidade a 100 °C e para os dados originais da viscosidade a -15 °C, os métodos só podem ser utilizados para controle de processos, enquanto que as demais propriedades podem ser empregadas tanto no controle de processos como em métodos laboratoriais.

#### 6.4.2 ESPECTRO COMPLETO – MIR

Analogamente aos modelos NIR, os modelos foram gerados utilizando-se todos os 794 comprimentos de onda dos espectros MIR para as 71 amostras de óleos lubrificantes comerciais. Da mesma forma, os espectros obtidos passaram por dois pré-tratamentos iniciais: os valores de transmitância foram transformados em absorbância e em seguida centrados na média.

Para as propriedades viscosidade a -15 °C e teor de magnésio, só foram fornecidos os valores referentes a 45 amostras, de forma que, para essas propriedades, os modelos PLS estão em uma matriz retangular contendo 45 linhas (amostras) e 794 colunas (comprimentos de onda). Devido ao pequeno número de amostras, optou-se pelo método de validação cruzada. Os resultados encontram-se listados nas TAB.s 6.9 e 6.10.

Para as demais propriedades, foram gerados modelos de calibração sem a presença das amostras testes, que foram separadas pelo método de ordenamento (TAB. 6.8) e utilizadas para a validação externa. Os resultados dos melhores modelos de predição com as amostras testes encontram-se nas TAB.s 6.11 a 6.13.

Na TAB. 6.14, estão apresentados os resultados obtidos pelo método de validação cruzada e ordenamento, mostrando-se, para cada propriedade, os valores de RMSEP obtidos para cada tipo de matriz espectral e o valor da reprodutibilidade do método de referência.

TAB. 6.8 Amostras Seleccionadas para o Conjunto Teste. Método Ordenamento – Validação Externa.

Propriedade	Amostras
Densidade	3,25,32,34,36,37,52,54,58,64
Ponto de fluidez	3,7,19,23,35,40,51,52,59,69
Viscosidade a 100°C	1,3,21,25,36,41,47,48,52,61

TAB. 6.9 PLS no MIR- Teor de Magnésio. Espectro Completo. Validação: Validação Cruzada.

Teor de Magnésio	Variáveis Latentes	Aberrantes	R	RMSEC
Original	10	12,20,21,36	0,9473	0,002
Autoescalados	6	20,33	0,8697	0,004
Normalizado	10	11,20,21	0,9540	0,003
Normalizado e Derivado	10	11,20,32,33	0,9606	0,002

TAB. 6.10 PLS no MIR- Viscosidade a -15 °C. Espectro Completo. Validação: Validação Cruzada.

Viscosidade a -15°C	Variáveis Latentes	Aberrantes	R	RMSEC
Original	5	4,7,10,13	0,9519	253
Autoescalados	6	4,7,13	0,9616	230
Normalizado	5	4,7,10,13	0,9539	247
Normalizado e Derivado	6	4,7,10,12	0,9606	234

TAB. 6.11 Modelo de Predição- Densidade. Espectro Completo  
Seleção de Amostras: Validação Externa.

Amostra	Medido	Normalizado	
		Predito	Erro
3	0.896	0.895	-0.0010
25	0.892	0.891	-0.0010
32	0.891	0.890	-0.0010
34	0.893	0.892	-0.0010
36	0.890	0.890	0.0000
37	0.889	0.890	0.0010
52	0.884	0.886	0.0020
54	0.882	0.887	0.0050
58	0.851	0.852	0.0010
64	0.890	0.885	-0.0050
PRESS			0.0001
RMSEP			0.0024



TAB. 6.12 Modelo de Predição- Ponto de Fluidez. Espectro Completo  
 Seleção de Amostras: Validação Externa.

Amostra	Medido	Normalizado	
		Predito	Erro
3	-9.000	-11.570	-2.57
7	-18.000	-13.381	4.62
19	-27.000	-24.621	2.38
23	-24.000	-25.917	-1.92
35	-24.000	-23.701	0.30
40	-27.000	-27.242	-0.24
51	-27.000	-26.966	0.03
52	-33.000	-26.277	6.72
59	-36.000	-39.692	-3.69
69	-30.000	-30.408	-0.41
PRESS			96.42
RMSEP			3.11

TAB. 6.13 Modelo de Predição- Viscosidade a 100°C. Espectro Completo  
 Seleção de Amostras: Validação Externa.

Amostra	Medido	Normalizado e Derivado	
		Predito	Erro
1	14.360	15.264	0.90
3	14.640	14.121	-0.52
21	20.770	20.049	-0.72
25	18.980	19.226	0.25
36	18.610	19.367	0.76
41	18.740	18.727	-0.01
47	15.350	15.928	0.58
48	14.870	15.709	0.83
52	19.620	17.965	-1.66
61	13.880	14.138	0.26
PRESS			6.08
RMSEP			0.78

TAB. 6.14 Resultados da Validação no MIR. Espectros Completos. Validação: Validação Cruzada e Ordenamento

Propriedades	RMSEP				Reproduz.
	Esp. Orig.	Esp. auto.	Esp. Norm.	Esp. NID.	
Densidade	0,0025 g/cm <sup>3</sup>	0,0030 g/cm <sup>3</sup>	0,0024 g/cm <sup>3</sup>	0,0029 g/cm <sup>3</sup>	0,0005g/cm <sup>3</sup>
Viscosidade a 100 °C	1,77 cSt	1,33 cSt	1,87 cSt	0,78 cSt	0,13 cSt
Viscosidade a -15 °C	256 cP	354 cP	319 cP	325 cP	670 cP
Ponto de Fluidez	3 °C	4 °C	3 °C	3 °C	6 °C
Teor de Magnésio	0,004 %(m)	0,006 %(m)	0,004 %(m)	0,004 %(m)	0,032 %(m)

#### 6.4.2.1 DISCUSSÃO – ESPECTRO COMPLETO (MIR)

Para a densidade, o menor valor de RMSEP foi observado para a matriz dos dados normalizados.

Já para a viscosidade a -15 °C, o menor valor de RMSEP foi obtido com os dados originais.

Para o teor de magnésio e ponto de fluidez, foram encontrados valores baixos de RMSEP para os dados originais, normalizados e para a primeira derivada dos dados normalizados.

Já para a viscosidade a 100 °C, o menor valor de RMSEP foi encontrado para a primeira derivada dos espectros normalizados.

Comparando-se os valores de RMSEP com os da reprodutibilidade, observa-se que, nos casos da densidade, viscosidade a 100 °C e para os dados autoescalados da viscosidade a -15 °C e do ponto de fluidez, os métodos só podem ser utilizados para controle de processos, enquanto que as demais propriedades podem ser empregadas tanto no controle de processos como em métodos laboratoriais.

#### 6.5 ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

##### 6.5.1 ANÁLISE DE COMPONENTES PRINCIPAIS – NIR

A análise por componentes principais (PCA) dos espectros no NIR é uma técnica que possibilita o exame da variação amostral apenas com as leituras espectrais das amostras, ou seja, não é necessária a medição de quaisquer propriedades físico-químicas.

O exame da variação das amostras é feito com o auxílio dos gráficos dos *loadings* para as primeiras componentes principais (PCs), onde as regiões do NIR cujos valores de *loadings* são altos em módulo, são as áreas mais importantes para a análise de regressão. Através da PCA, também é possível realizar a descrição da variação de dados multivariados através de um pequeno número de variáveis latentes ortogonais (não correlacionadas) entre si, obtidas através de combinações lineares das variáveis.



Para uma melhor compreensão da variação das amostras e para a seleção dos comprimentos de onda (variáveis independentes) que contivessem informações relevantes, usadas na análise de regressão, foi efetuada uma análise por componentes principais na matriz contendo os espectros no NIR das 60 amostras de óleos lubrificantes comerciais. A matriz dos dados é retangular, contendo 60 linhas (amostras) e 680 colunas (comprimentos de onda entre 800 e 2240 nm). Os espectros obtidos passaram por dois pré-tratamentos iniciais: os valores de transmitância foram transformados em absorbância e em seguida centrados na média.

No gráfico da variância explicada contra componentes principais, que se encontra na FIG. 6.12, pode-se observar que os três primeiros PCs explicam 95% da variância original contida nas 680 variáveis originais, enquanto que os seis primeiros respondem por 99% da variação contida nos dados originais.

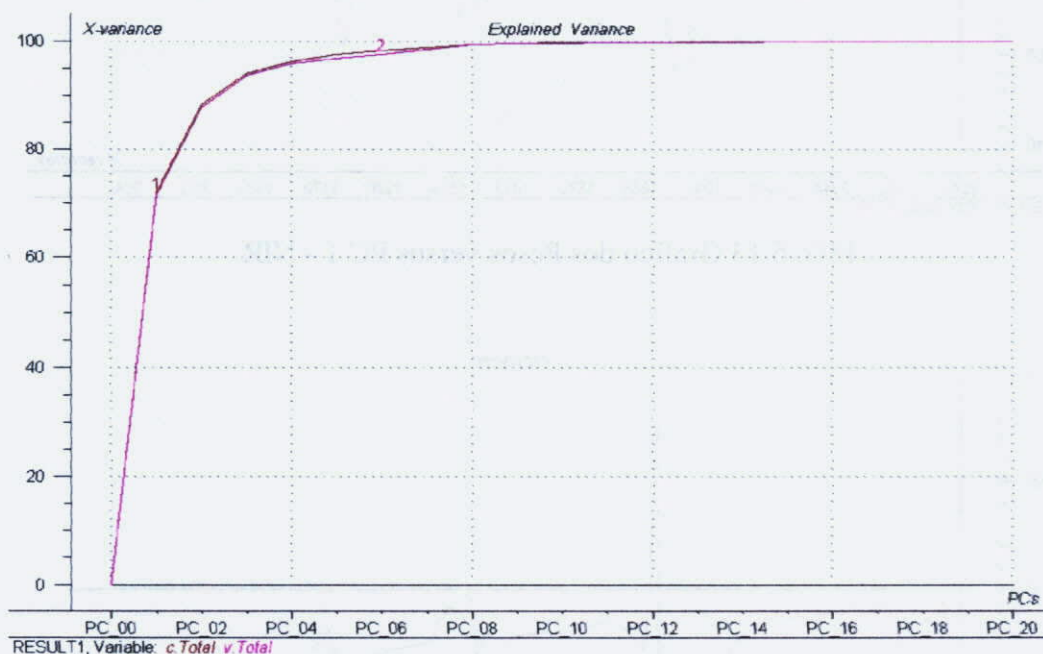


FIG.6.12 Gráfico da Variância explicada versus Componentes principais (PCs) – NIR

Nas FIG.s 6.13 a 6.15 encontram-se os gráficos dos *loadings* (pesos) contra as três primeiras PCs. A análise destes gráficos permite verificar que os comprimentos de onda que apresentaram os maiores pesos nos três primeiros e mais importantes PCs correspondem às absorções entre 1620 e 1780 nm (devido a sobreposição de bandas geradas pelas vibrações de primeiro sobretom de ligações C-H olefinicas, parafinicas e aromáticas e das vibrações de deformação axial C=O dos ésteres) e entre 2110 e 2240 nm (combinação da vibração axial fundamental com a vibração angular fundamental de ligações C-H aromáticas). Observa-se também a presença de uma banda de média intensidade ao redor de 1210 nm (referentes às

vibrações axiais de segundo sobretom de ligações C-H) e de dois picos de fraca intensidade entre 1390 e 1420 nm.

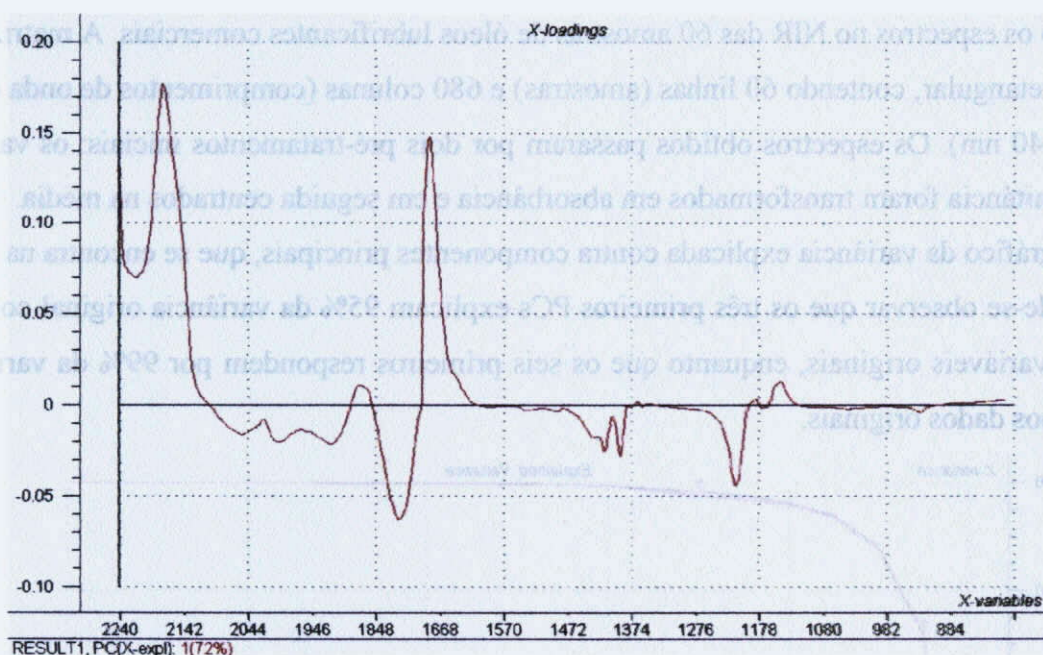


FIG. 6.13 Gráfico dos Pesos versus PC 1 - NIR

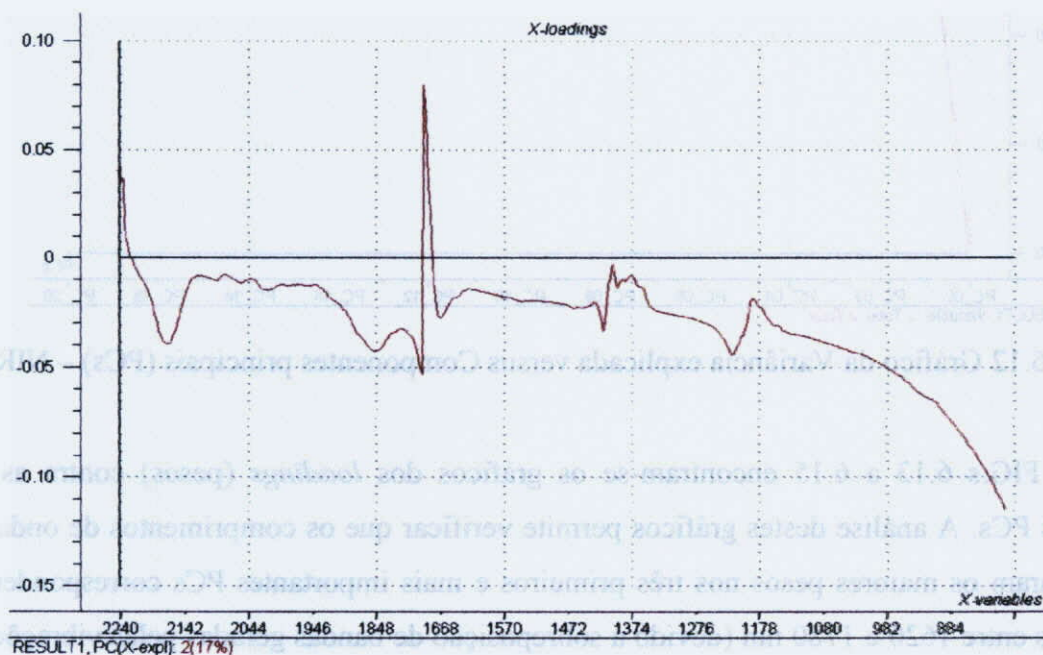


FIG. 6.14 Gráfico dos Pesos versus PC 2 - NIR





FIG. 6.15 Gráfico dos Pesos versus PC 3 – NIR

A análise de amostras aberrantes foi realizada através da FIG. 6.16, gráfico de variância residual versus os valores de *leverage*, como descrito no item 4.3.1.2.b.

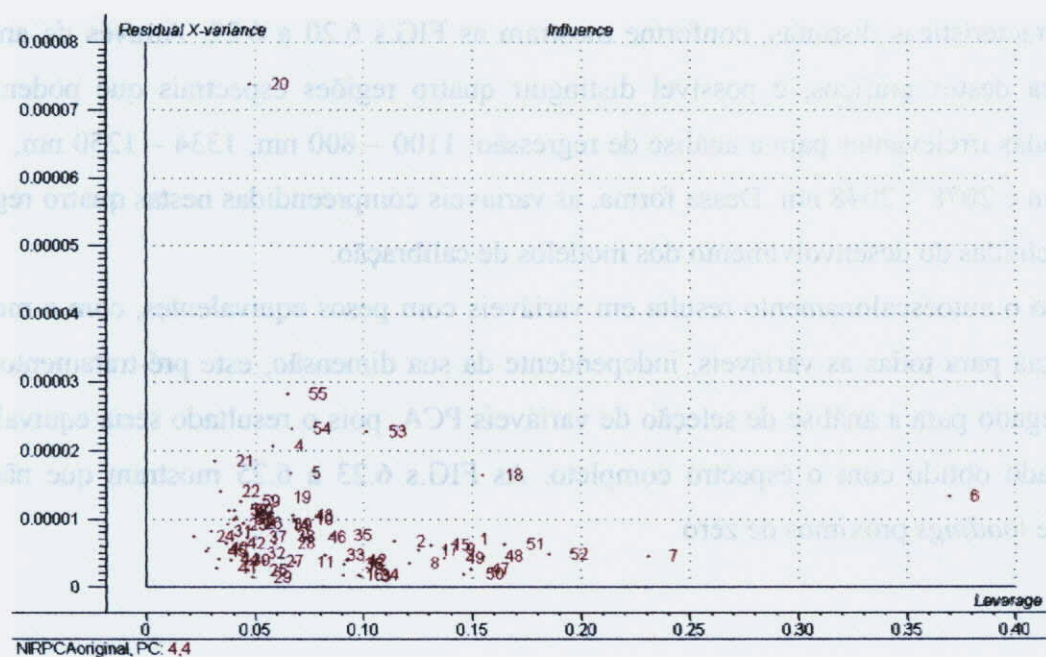


FIG. 6.16 Gráfico do Resíduo versus *Leverage* – NIR

Do gráfico 6.16, é possível concluir que, das amostras estudadas, nenhuma apresenta simultaneamente altos valores de *leverage* e resíduo. Entretanto, como a média dos valores de

*leverage* está em torno de 0,1, e como a amostra 6 apresenta um valor de *leverage* maior do que o triplo deste valor médio, esta deve ser observada com atenção durante o desenvolvimento dos modelos de calibração.

#### 6.5.1.1 SELEÇÃO DE VARIÁVEIS - NIR

Foram efetuadas novas análises por componentes principais na matriz dos dados originais NIR, normalizados e na primeira derivada dos espectros normalizados. O resultado da PCA para os espectros originais foi obtido através da análise simultânea dos gráficos dos pesos 6.13 a 6.15 do item 6.5.1. Pode-se observar que não é possível a identificação simultânea de uma ou mais regiões dos espectros cujos comprimentos de onda não sejam relevantes, em outras palavras, que apresentam pesos iguais ou muito próximos de zero. Sendo assim, nenhum comprimento de onda poder ser eliminado da análise de regressão. O mesmo pode ser dito para os espectros normalizados, conforme as FIG.s 6.17 a 6.19.

Entretanto, o exame da PCA aplicada aos espectros normalizados e então derivados revela características distintas, conforme mostram as FIG.s 6.20 a 6.22. Através da análise simultânea destes gráficos, é possível distinguir quatro regiões espectrais que podem ser consideradas irrelevantes para a análise de regressão: 1100 – 800 nm, 1334 – 1250 nm, 1612 – 1550 nm e 2078 – 2048 nm. Dessa forma, as variáveis compreendidas nestas quatro regiões foram excluídas do desenvolvimento dos modelos de calibração.

Como o autoescalonamento resulta em variáveis com pesos equivalentes, com a mesma importância para todas as variáveis, independente da sua dimensão, este pré-tratamento não foi empregado para a análise de seleção de variáveis PCA, pois o resultado seria equivalente ao resultado obtido com o espectro completo. As FIG.s 6.23 a 6.25 mostram que não há regiões de *loadings* próximos de zero.



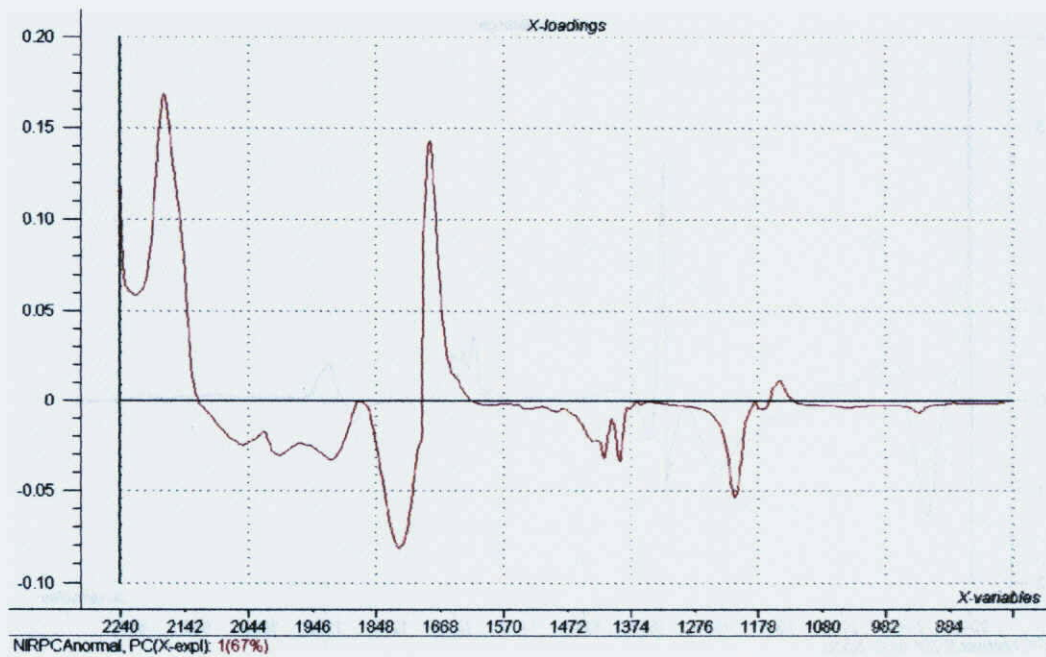


FIG. 6.17 Gráfico dos Pesos versus PC 1 – NIR: espectros normalizados

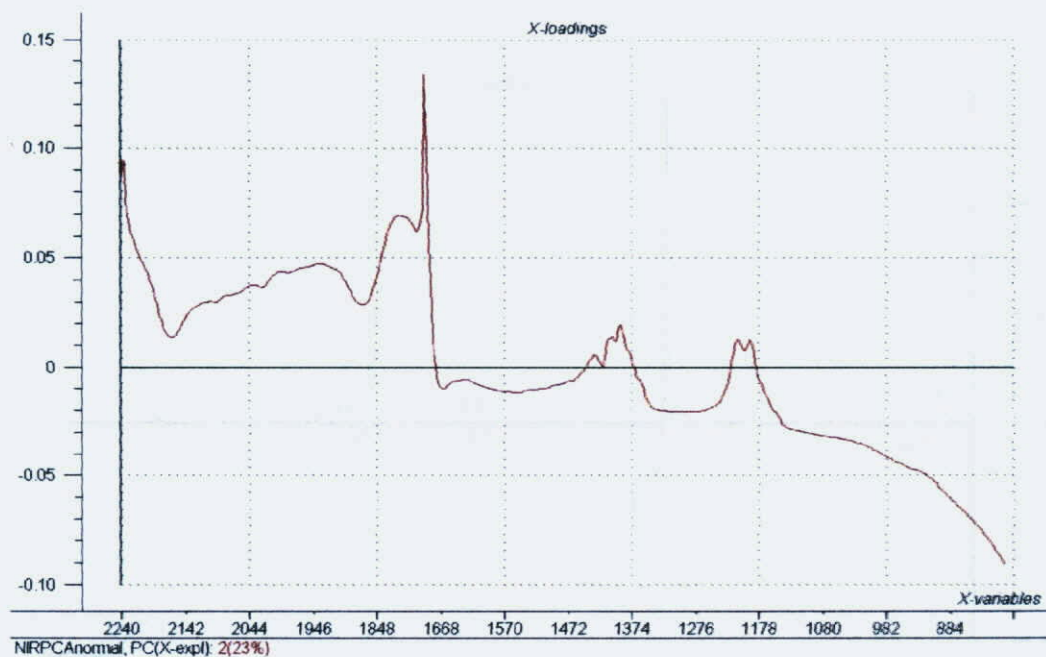


FIG. 6.18 Gráfico dos Pesos versus PC 2 – NIR: espectros normalizados

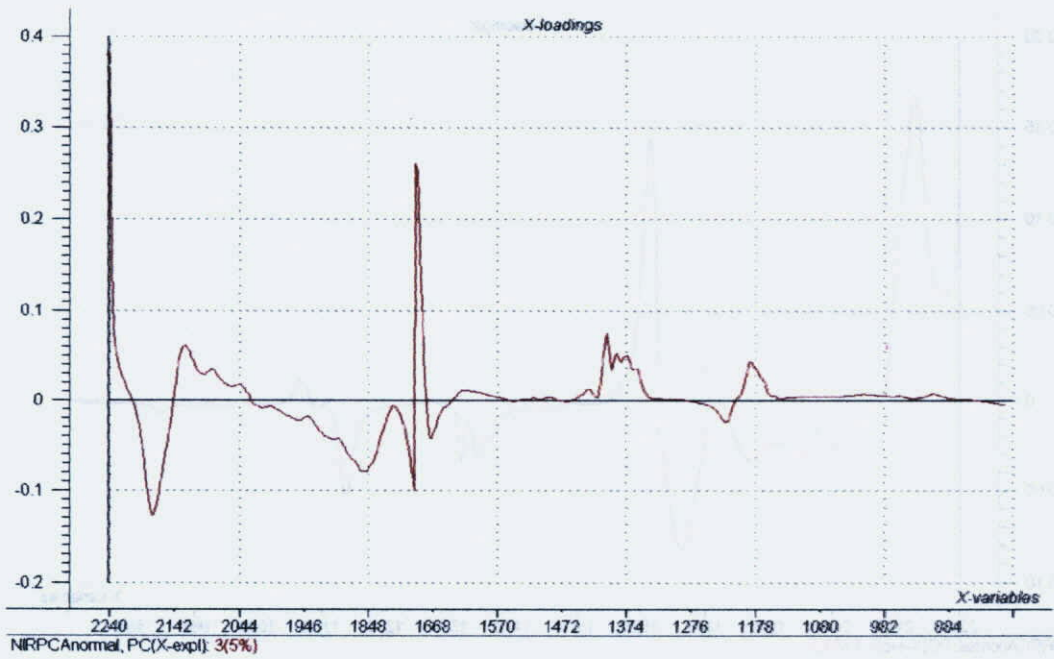


FIG. 6.19 Gráfico dos Pesos versus PC 3 – NIR: espectros normalizados



FIG. 6.20 Gráfico dos Pesos versus PC 1 – NIR: espectros normalizados e derivados



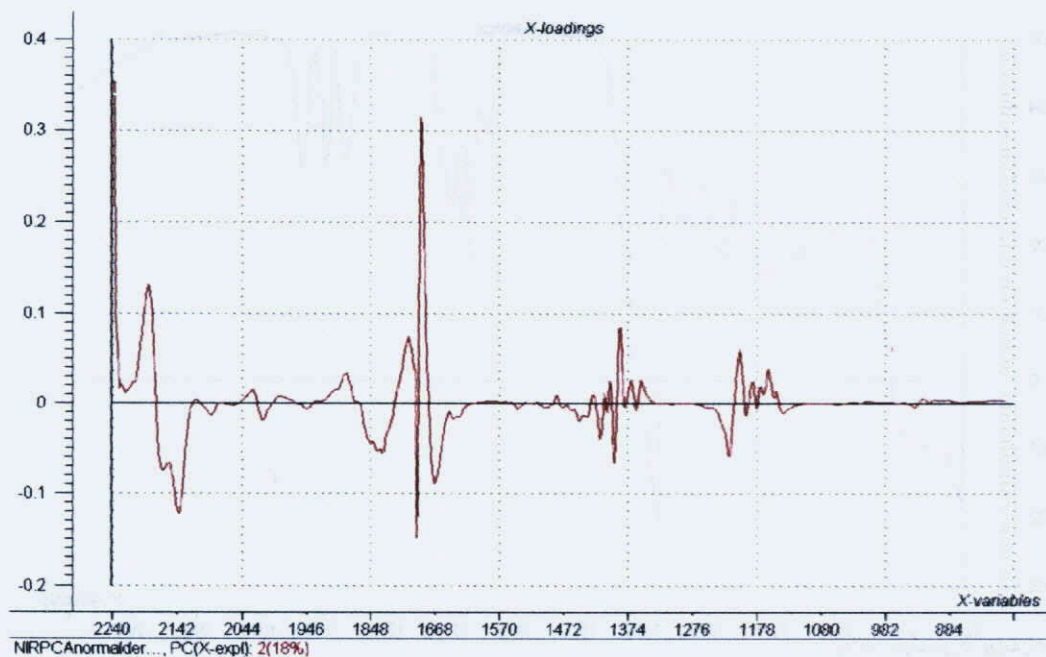


FIG. 6.21 Gráfico dos Pesos versus PC 2 – NIR: espectros normalizados e derivados

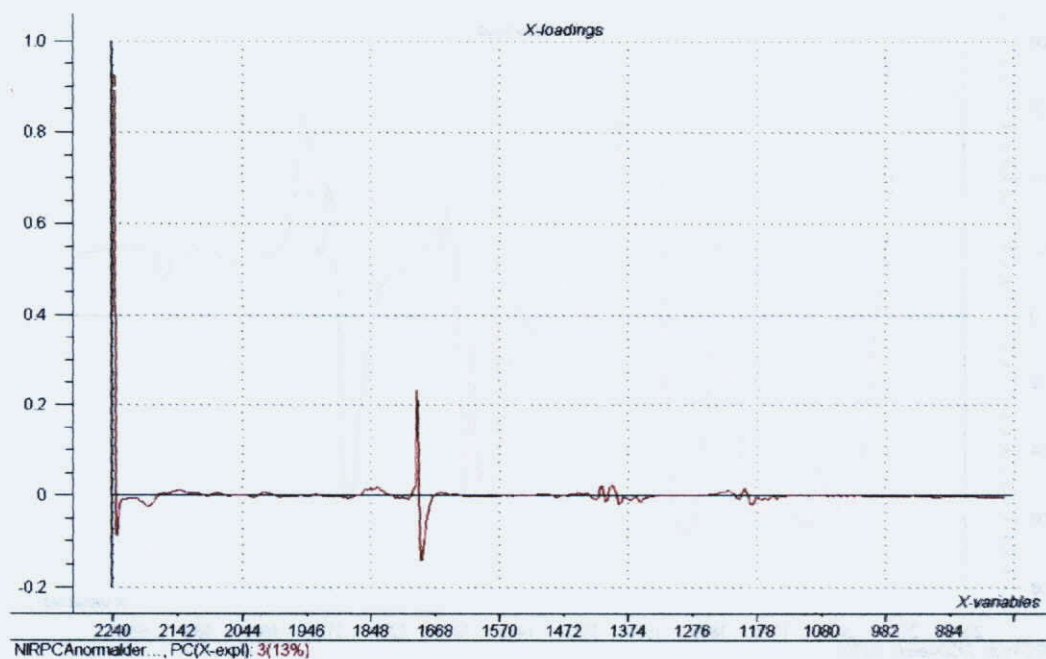


FIG. 6.22 Gráfico dos Pesos versus PC 3 – NIR: espectros normalizados e derivados

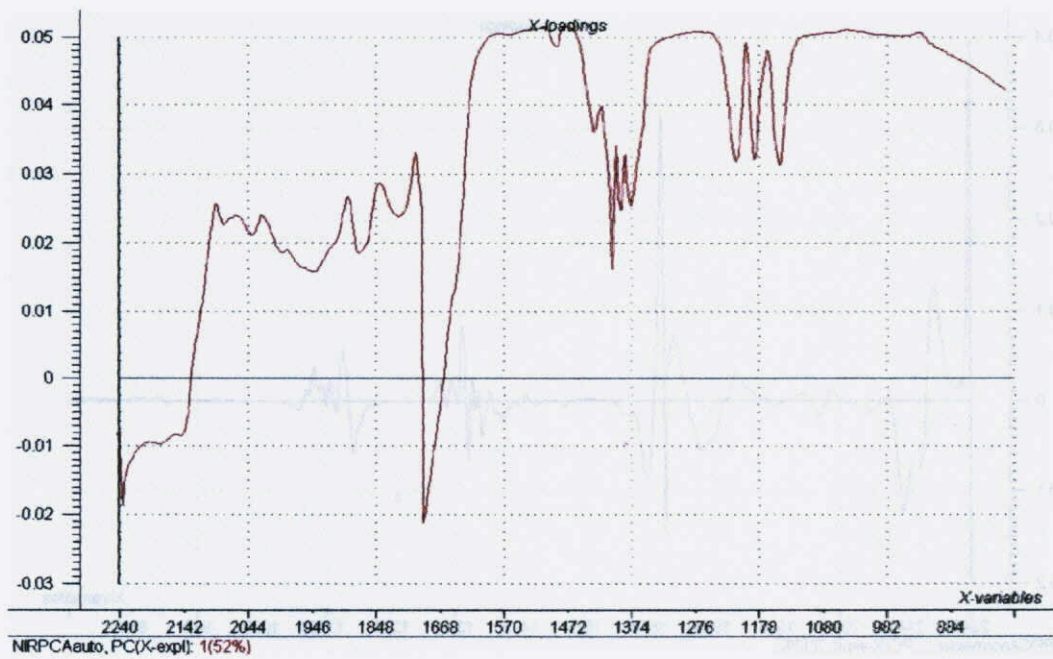


FIG. 6.23 Gráfico dos Pesos versus PC 1 – NIR: espectros autoescalados

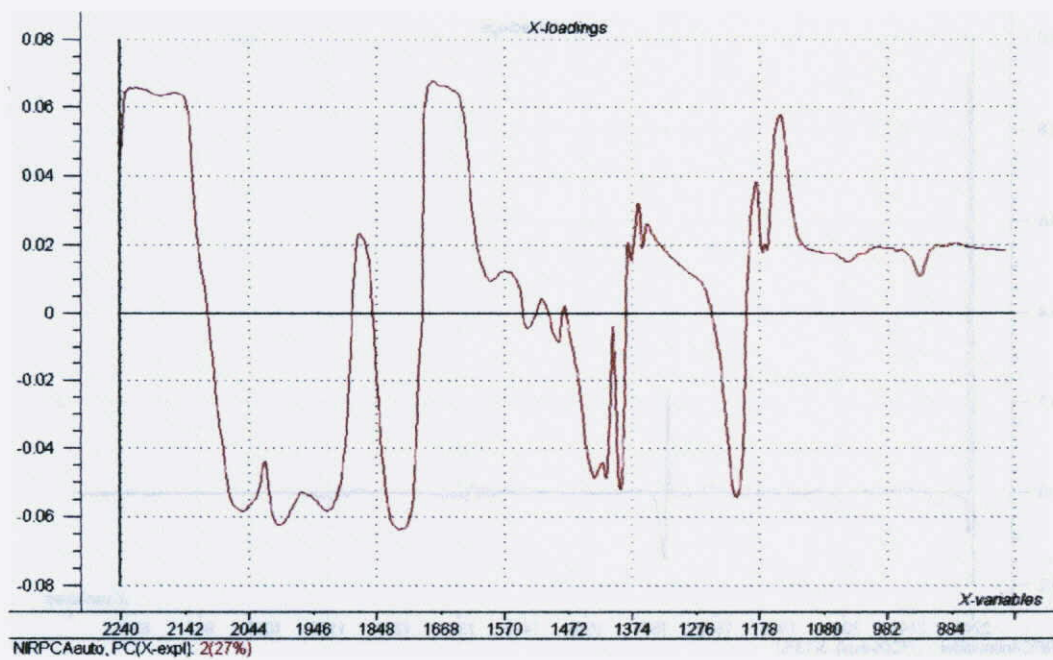


FIG. 6.24 Gráfico dos Pesos versus PC 2 – NIR: espectros autoescalados



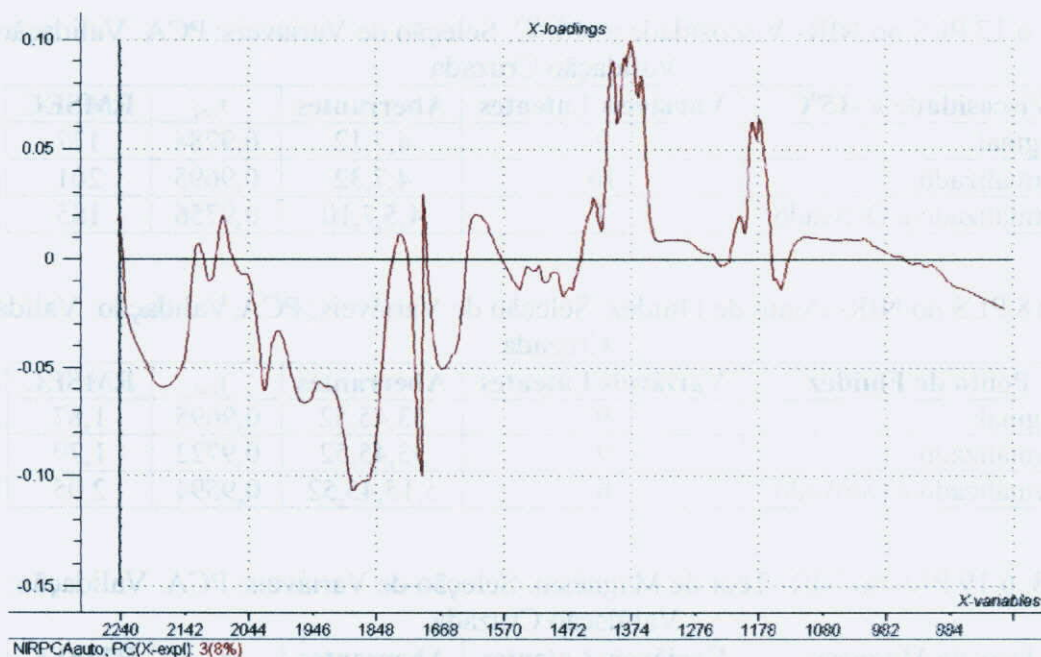


FIG. 6.25 Gráfico dos Pesos versus PC 3 – NIR: espectros autoescalados

Nas TAB.s 6.15 a 6.19, encontram-se os resultados da calibração obtidos para as cinco propriedades estudadas, da mesma forma que na Seção 6.4.1 (TAB.s 6.2 a 6.6). Cabe lembrar que para as propriedades viscosidade a  $-15\text{ }^{\circ}\text{C}$  e teor de magnésio utilizaram-se 38 amostras, enquanto que para as demais propriedades usou-se 60 amostras.

Na TAB. 6.20 estão apresentados os resultados obtidos pelo método da validação cruzada. Nessa tabela estão os valores de RMSEP para cada tipo de matriz espectral obtidos em cada propriedade, bem como o valor da reprodutibilidade do método de referência.

TAB. 6.15 PLS no NIR- Densidade. Seleção de Variáveis: PCA. Validação: Validação Cruzada.

Densidade	Variáveis Latentes	Aberrantes	$r_{cal}$	RMSEC
Original	4	18,27,38	0,9853	0,0021
Normalizado	5	18,27,55	0,9911	0,0017
Normalizado e Derivado	5	27	0,9916	0,0016

TAB. 6.16 PLS no NIR- Viscosidade a  $100\text{ }^{\circ}\text{C}$ . Seleção de Variáveis: PCA. Validação: Validação Cruzada.

Viscosidade a $100\text{ }^{\circ}\text{C}$	Variáveis Latentes	Aberrantes	$r_{cal}$	RMSEC
Original	11	9,20,46	0,9881	0,39
Normalizado	10	9	0,9798	0,52
Normalizado e Derivado	10	9,20,36	0,9859	0,43

TAB. 6.17 PLS no NIR- Viscosidade a -15 °C. Seleção de Variáveis: PCA .Validação: Validação Cruzada.

Viscosidade a -15°C	Variáveis Latentes	Aberrantes	r <sub>cal</sub>	RMSEC
Original	9	4,7,12	0,9784	177
Normalizado	10	4,7,32	0,9695	201
Normalizado e Derivado	7	4,5,7,10	0,9756	185

TAB. 6.18 PLS no NIR- Ponto de Fluidez. Seleção de Variáveis: PCA. Validação: Validação Cruzada.

Ponto de Fluidez	Variáveis Latentes	Aberrantes	r <sub>cal</sub>	RMSEC
Original	9	33,45,52	0,9695	1,87
Normalizado	9	33,45,52	0,9722	1,79
Normalizado e Derivado	6	5,15,45,52	0,9594	2,05

TAB. 6.19 PLS no NIR- Teor de Magnésio. Seleção de Variáveis: PCA. Validação: Validação Cruzada.

Teor de Magnésio	Variáveis Latentes	Aberrantes	r <sub>cal</sub>	RMSEC
Original	4	25	0,7709	0,005
Normalizado	5	12,21,25,28	0,8431	0,003
Normalizado e Derivado	8	10,20,27,31	0,9457	0,003

TAB. 6.20 Resultados da Validação no NIR. Seleção de variáveis: PCA. Validação : Validação Cruzada

Propriedades	RMSEP			Reproduz.
	Esp. Orig.	Esp. Norm.	Esp. N1D.	
Densidade	0,0026 g/cm <sup>3</sup>	0,0020 g/cm <sup>3</sup>	0,0020 g/cm <sup>3</sup>	0,0005g/cm <sup>3</sup>
Viscosidade a 100 °C	0,61 cSt	0,78 cSt	0,67 cSt	0,13 cSt
Viscosidade a -15 °C	339 cP	386 cP	337 cP	670 cP
Ponto de Fluidez	3 °C	3 °C	3 °C	6 °C
Teor de Magnésio	0,007 %(m)	0,005 %(m)	0,006 %(m)	0,032 %(m)

#### 6.5.1.2 DISCUSSÃO - ANÁLISE DE COMPONENTES PRINCIPAIS (NIR)

Para a densidade, os melhores resultados foram observados com a normalização e a primeira derivada dos espectros normalizados.

Para a viscosidade a 100 °C, observou-se que o melhor resultado foi obtido sem nenhum pré-tratamento.

Para o teor de magnésio, o menor valor de RMSEP foi observado para a normalização.



Para a viscosidade a  $-15\text{ }^{\circ}\text{C}$ , o melhor resultado de RMSEP provou ser a primeira derivada dos espectros normalizados.

Para o ponto de fluidez, todas as matrizes utilizadas apresentam resultados de RMSEP igualmente satisfatórios.

Comparando-se o método de seleção de variáveis PCA com os espectros completos pode-se observar que:

- Densidade: os resultados encontrados foram muito semelhantes, e os menores valores de RMSEP foram novamente obtidos nas matrizes que passaram por algum pré-tratamento.
- Viscosidade a  $100\text{ }^{\circ}\text{C}$ : os dados obtidos com a seleção de variáveis PCA apresentaram uma pequena diminuição nos valores de RMSEP em relação ao espectro completo para os dados originais.
- Viscosidade a  $-15\text{ }^{\circ}\text{C}$ : a utilização da seleção de variáveis PCA aumentou os valores de RMSEP para as matrizes com pré-tratamento.
- Teor de magnésio: somente nos dados originais, para a seleção de variáveis PCA, o resultado de RMSEP foi pior do que os valores para a utilização do espectro completo.
- Ponto de fluidez: os valores de RMSEP para a seleção de variáveis PCA e espectro completos foram praticamente iguais para as matrizes com pré-tratamento, obtendo-se uma pequena diminuição no valor de RMSEP para a matriz dos dados originais dos espectros completos.

Comparando-se os valores de RMSEP com a reprodutibilidade, pode-se perceber que, novamente, a densidade e a viscosidade a  $100\text{ }^{\circ}\text{C}$  só podem ser utilizados para controle de processo. Neste caso, a matriz dos dados normalizados apresentou um valor de RMSEP maior que duas vezes à reprodutibilidade para a viscosidade a  $-15\text{ }^{\circ}\text{C}$ , diferente do método que utiliza o espectro completo, onde a matriz dos dados originais apresentou um valor superior a duas vezes a reprodutibilidade.

Para o teor de magnésio, todas as matrizes empregadas apresentaram resultados inferiores a duas vezes o valor da reprodutibilidade. Enquanto que para o ponto de fluidez todas as matrizes apresentaram resultados aproximadamente iguais a duas vezes o valor da reprodutibilidade.

## 6.5.2 ANÁLISE DE COMPONENTES PRINCIPAIS – MIR

Foi realizada uma análise por componente principal na matriz contendo os espectros no MIR das setenta e uma amostras de óleos lubrificantes comerciais. A matriz dos dados é retangular, contendo 71 linhas (amostras) e 794 colunas (comprimentos de onda entre 2000-400  $\text{cm}^{-1}$ ). Os valores sofreram dois pré-tratamentos: os valores de transmitância foram transformados em absorbância e em seguida centrados na média.

O gráfico da variância explicada versus PCs encontra-se na FIG. 6.26. Observa-se que os cinco primeiros PCs explicam 98% da variância original e que os seis primeiros respondem por 99% da variância contida nos dados originais.

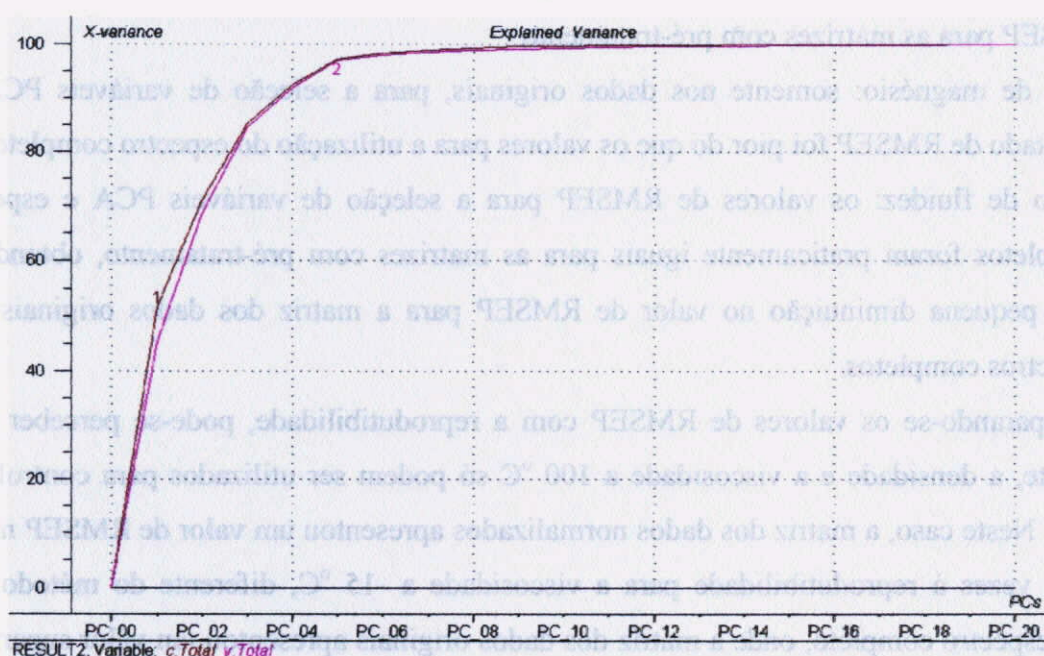


FIG. 6.26 Gráfico de variância explicada versus componentes principais - MIR

Nas FIG. 6.27 a 6.30, encontram-se os gráficos dos pesos versus os quatro primeiros PCs. A análise de tais gráficos revela os comprimentos de onda que apresentaram os maiores pesos nos quatro primeiros e mais importantes PCs, correspondentes às vibrações fundamentais do tipo rock de ligações C-H, onde o número de átomos de carbono em cadeia é maior ou igual á quatro (ao redor de 721  $\text{cm}^{-1}$ ), às vibrações fundamentais angulares de ligações C-H (entre 1350-1450  $\text{cm}^{-1}$ ) e às vibrações de deformações axial C=O (entre 1776-1735  $\text{cm}^{-1}$ ) e C-O (entre 1300-1000  $\text{cm}^{-1}$ ) dos ésteres.



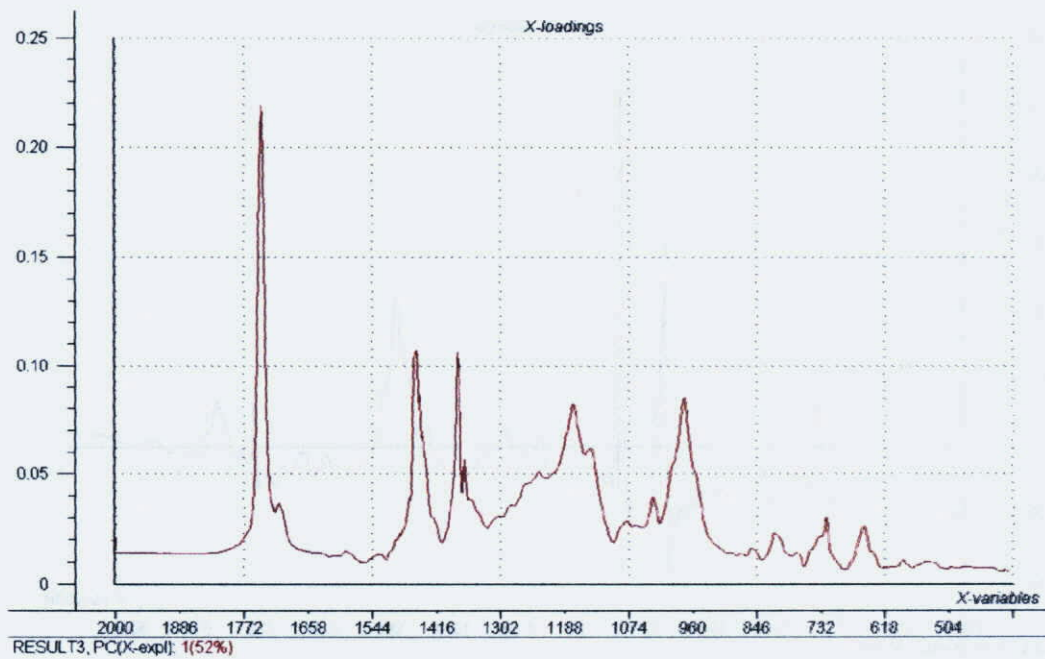


FIG. 6.27 Gráfico dos Pesos versus componentes principais 1 - MIR

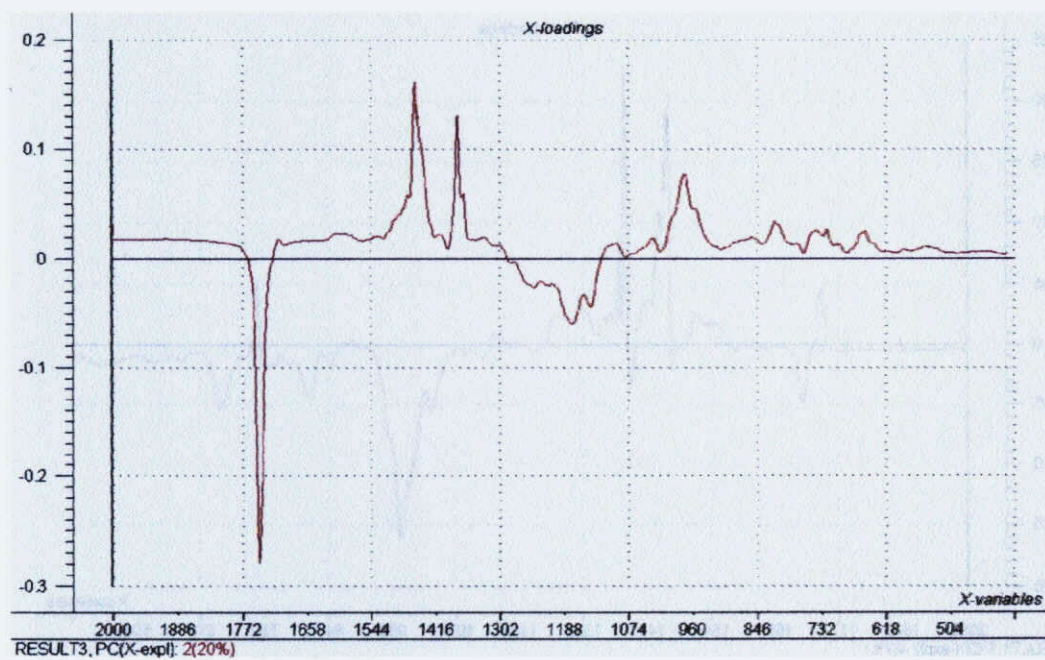


FIG. 6.28 Gráfico dos Pesos versus componentes principais 2 - MIR

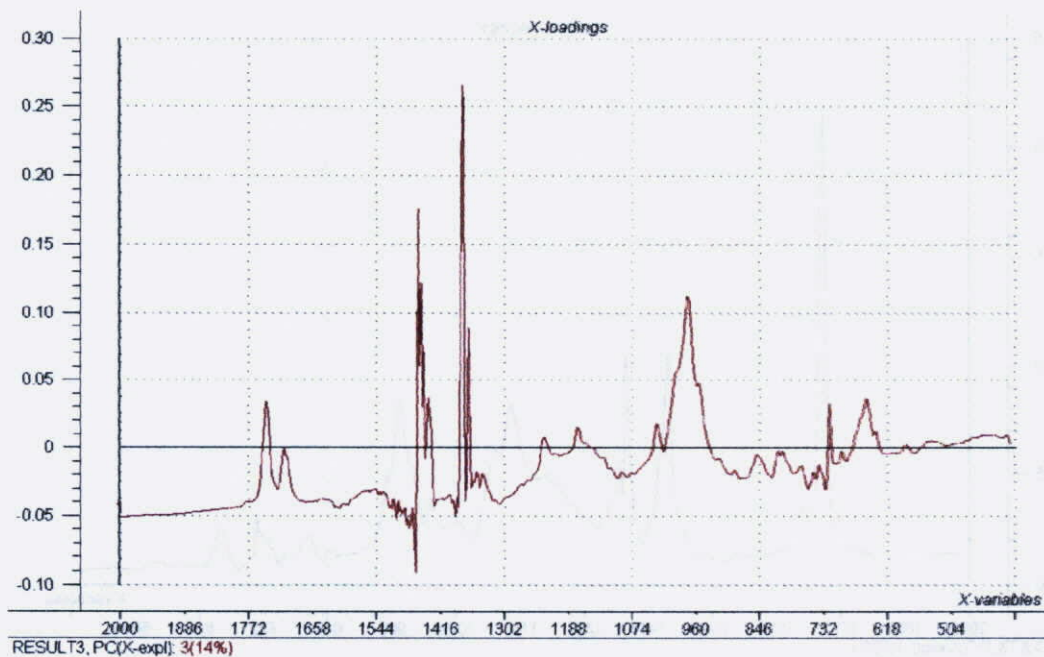


FIG. 6.29 Gráfico dos Pesos versus componentes principais 3 – MIR

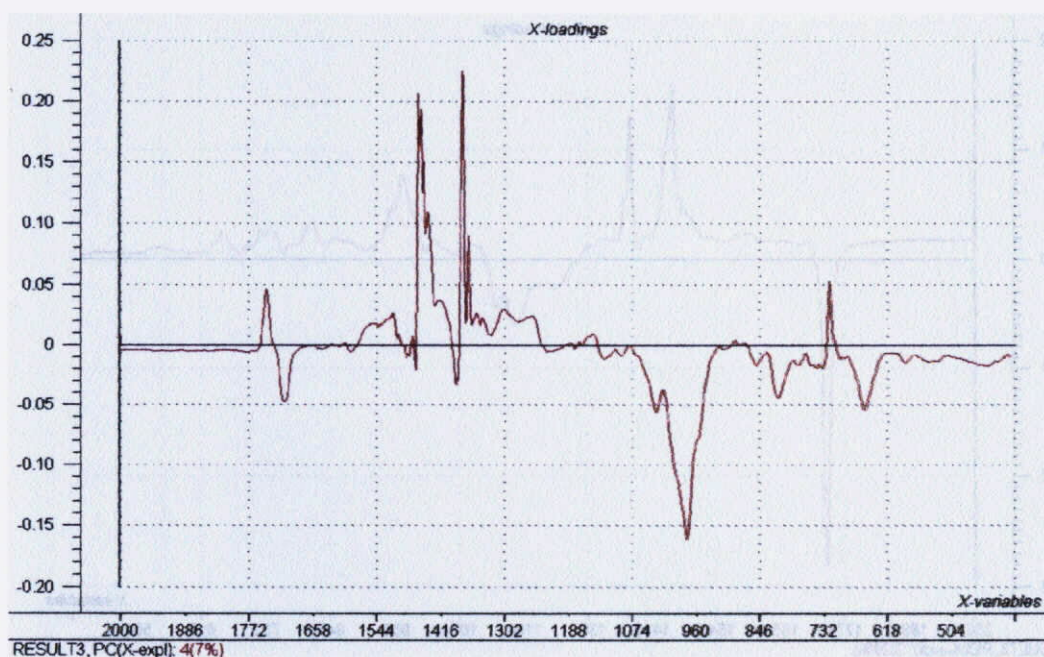


FIG. 6.30 Gráfico dos Pesos versus componentes principais 4 - MIR

Foi construído um gráfico de variância residual versus *leverage* para a identificação de amostras aberrantes (FIG. 6.31). Deste gráfico, é possível concluir que as amostras de 64 a 66 apresentam um valor superior a três vezes o valor médio do *leverage*. Vale salientar que estas são as amostras sintéticas da BARDAHL, que possuem propriedades distintas em relação às demais. A amostra 5, embora possua um alto valor residual, não apresenta um alto valor de



*leverage*. Sendo assim, as amostras citadas acima não devem ser eliminadas, e sim avaliadas com atenção durante o desenvolvimento dos modelos de regressão.

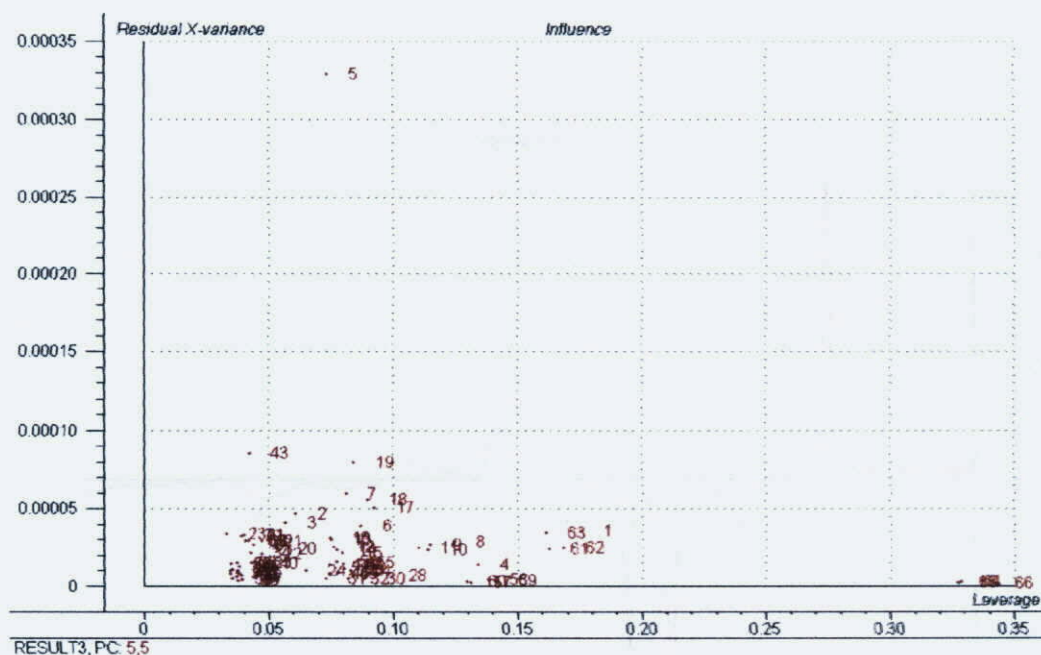


FIG. 6.31 Gráfico do Resíduo versus *Leverage* – MIR

### 6.5.2.1 SELEÇÃO DE VARIÁVEIS - MIR

Foram efetuadas novas análises de componentes principais, agora para a matriz MIR dos dados originais, com autoescalonamento, normalizados e para a primeira derivada dos espectros normalizados. A PCA para os espectros originais foi feita pela análise simultânea dos gráficos dos pesos 6.27 á 6.30 do item 6.3.2. Pode-se observar que nenhum comprimento de onda pode ser eliminado da análise de regressão. O mesmo pode ser dito para os espectros normalizados, conforme as FIG.s 6.32 a 6.35.

O exame da PCA aplicada à primeira derivada dos espectros normalizados revela características distintas, conforme mostram as FIG.s 6.36 a 6.39. Podem-se distinguir, através das análises simultâneas destes gráficos, três regiões espectrais que podem ser consideradas irrelevantes para a análise de regressão: 514 – 400 nm, 1686 – 1632 nm e 2000 – 1768 nm. As variáveis compreendidas nestas três regiões foram excluídas para o desenvolvimento do modelo de calibração.

Como o autoescalonamento atribui a mesma importância para todas as variáveis, podemos observar que as FIG.s 6.40 a 6.43 não mostram regiões de *loadings* próximos de zero.

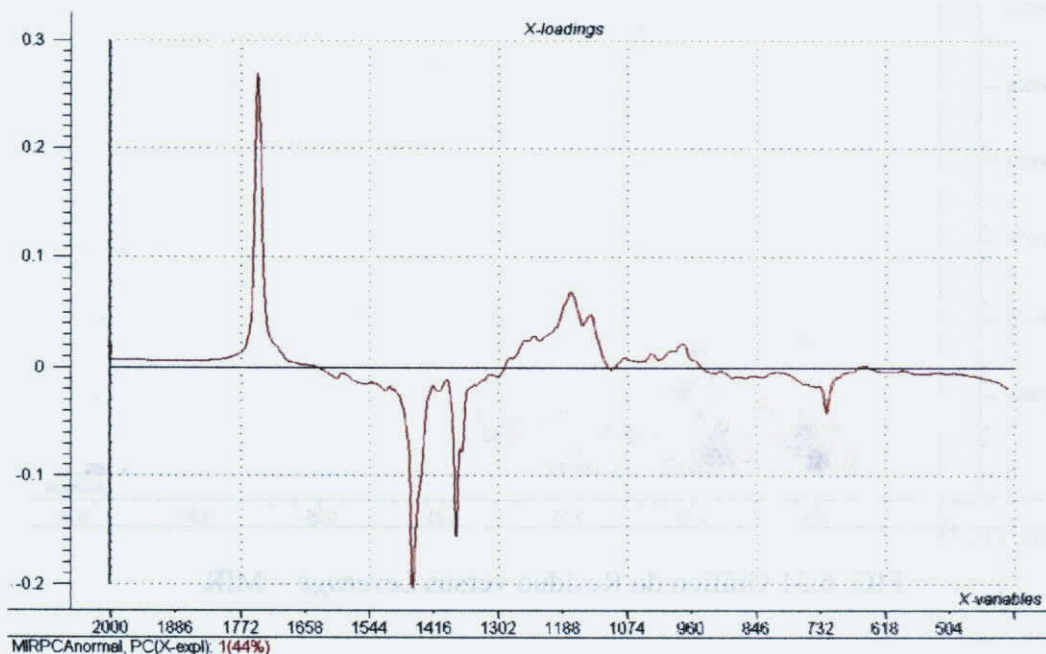


FIG. 6.32 Gráfico dos Pesos versus PC 1 – MIR: espectros normalizados

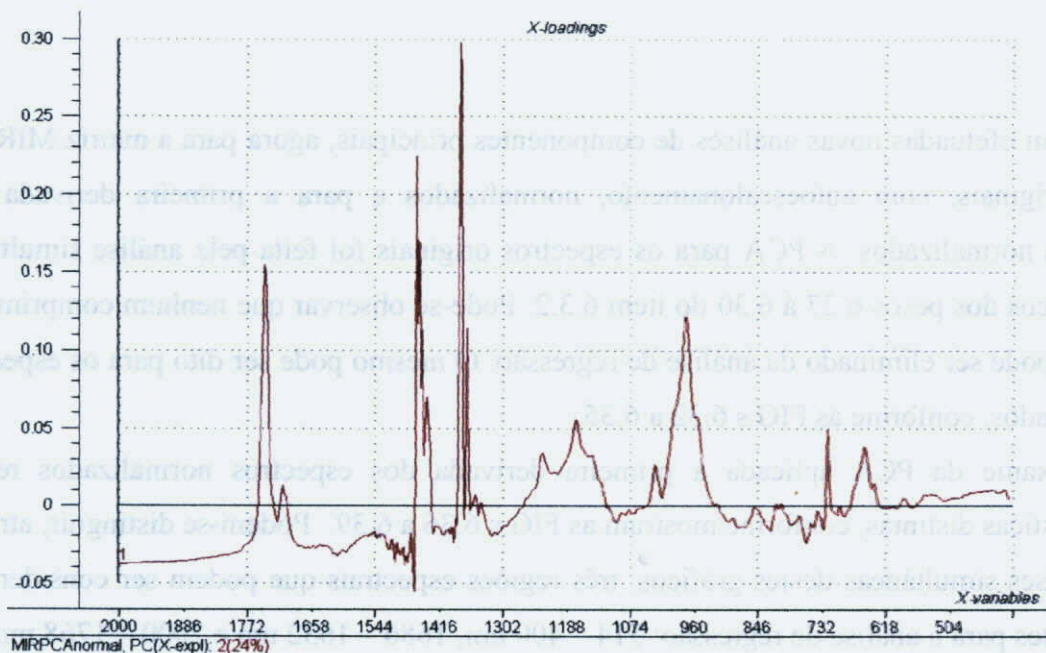


FIG. 6.33 Gráfico dos Pesos versus PC 2 – MIR: espectros normalizados



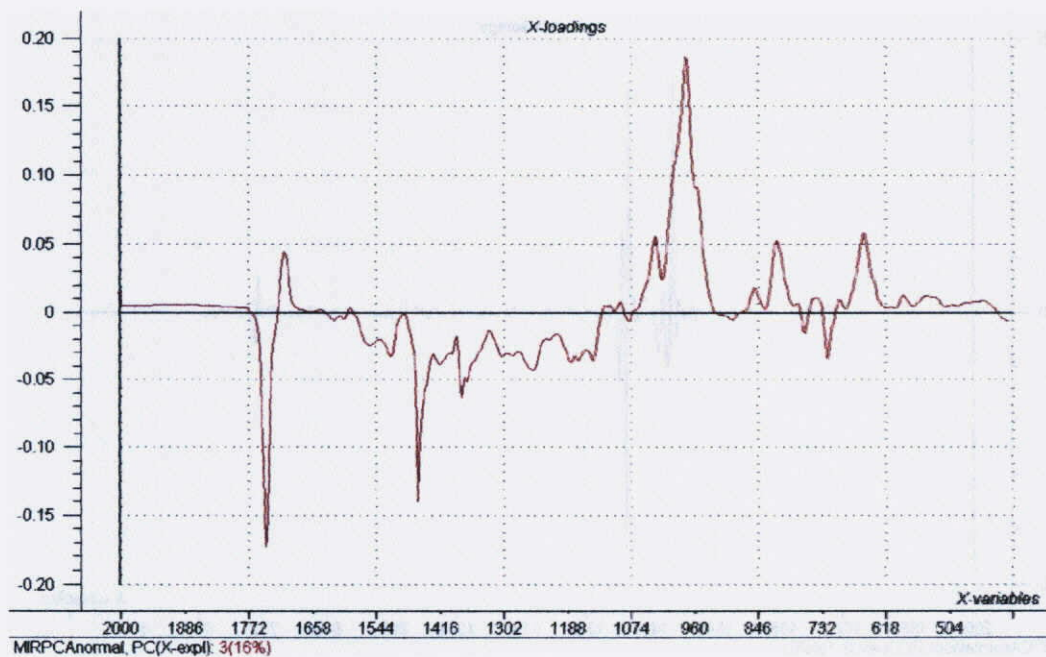


FIG. 6.34 Gráfico dos Pesos versus PC 3 – MIR: espectros normalizados

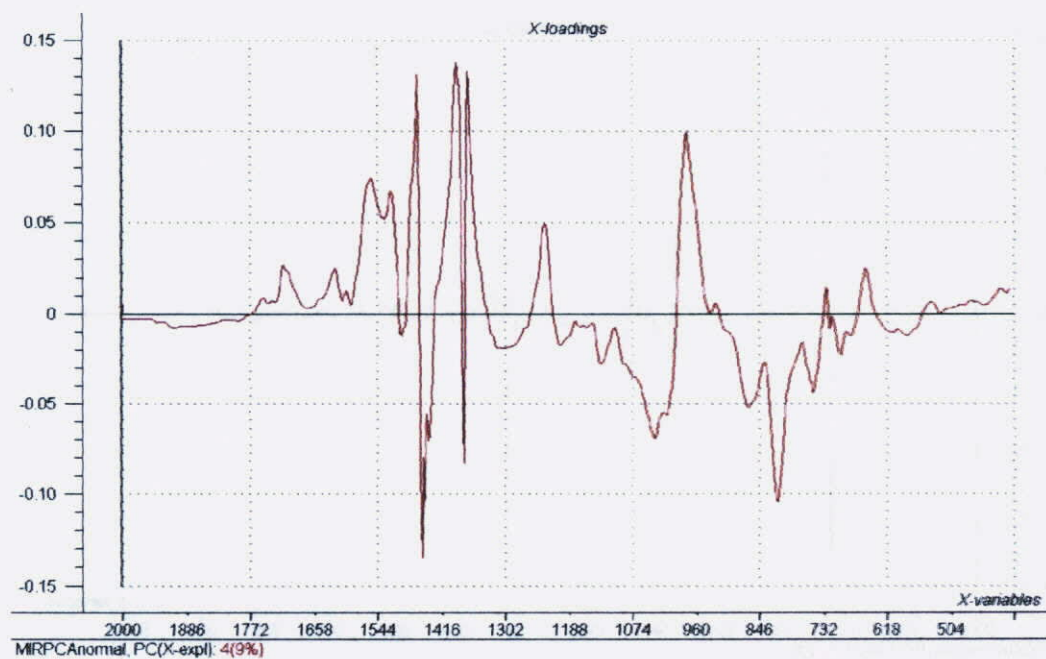


FIG. 6.35 Gráfico dos Pesos versus PC 4 – MIR: espectros normalizados

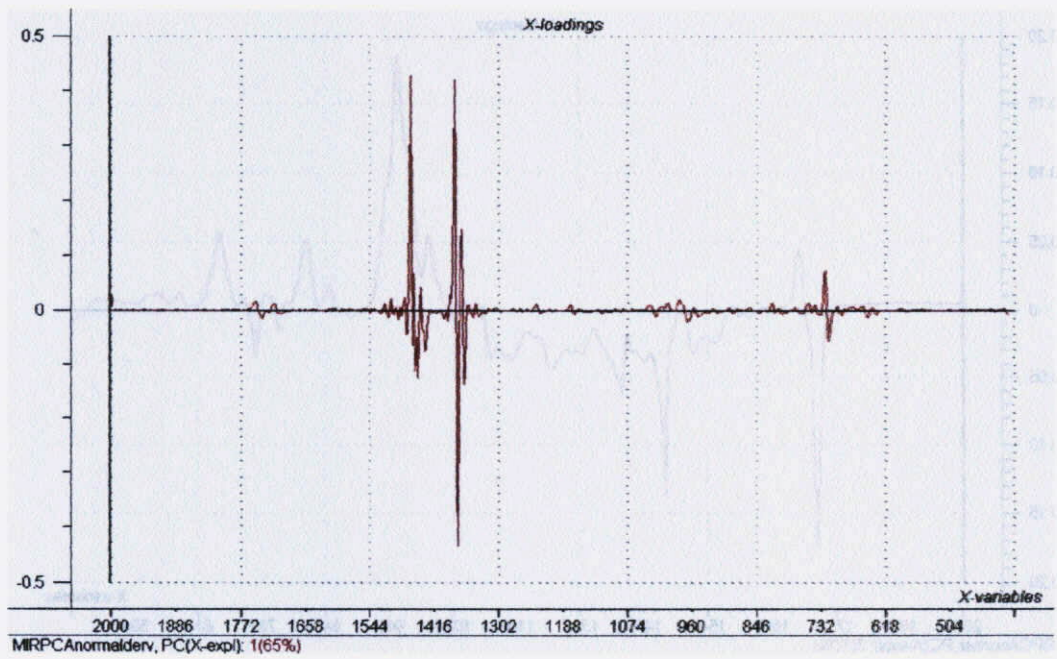


FIG. 6.36 Gráfico dos Pesos versus PC 1 – MIR: espectros normalizados e derivados

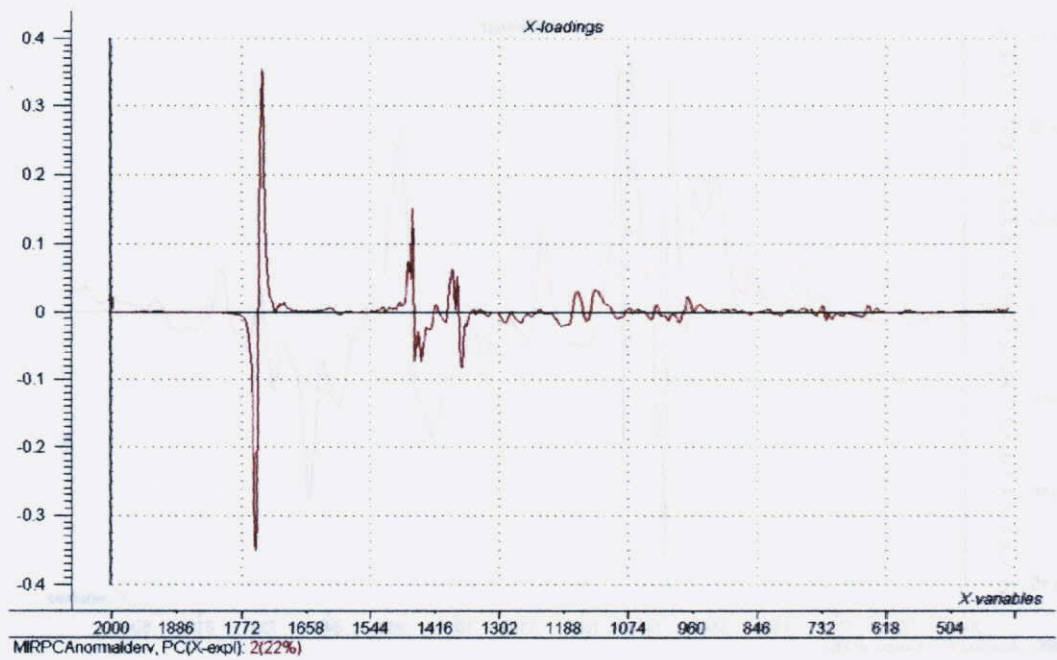


FIG. 6.37 Gráfico dos Pesos versus PC 2 – MIR: espectros normalizados e derivados



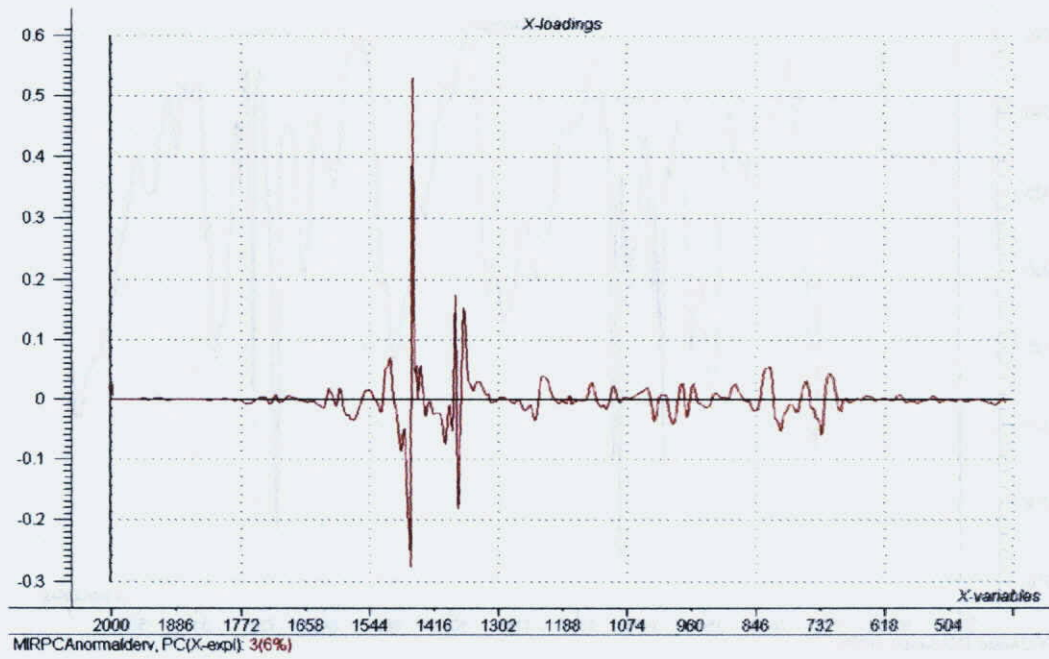


FIG. 6.38 Gráfico dos Pesos versus PC 3 – MIR: espectros normalizados e derivados

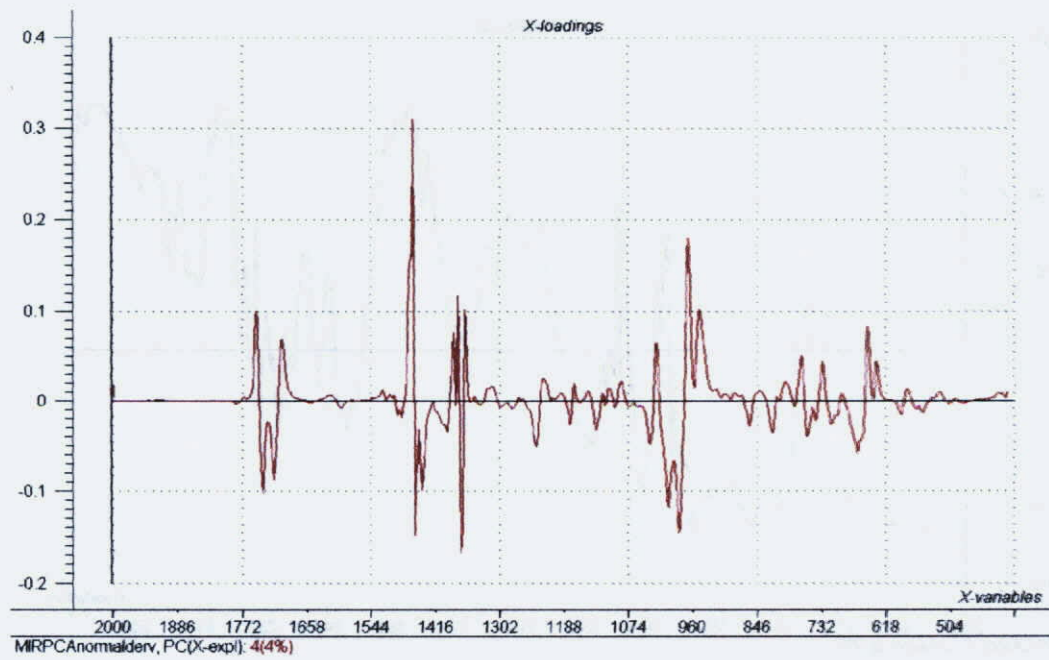


FIG. 6.39 Gráfico dos Pesos versus PC 4 – MIR: espectros normalizados e derivados

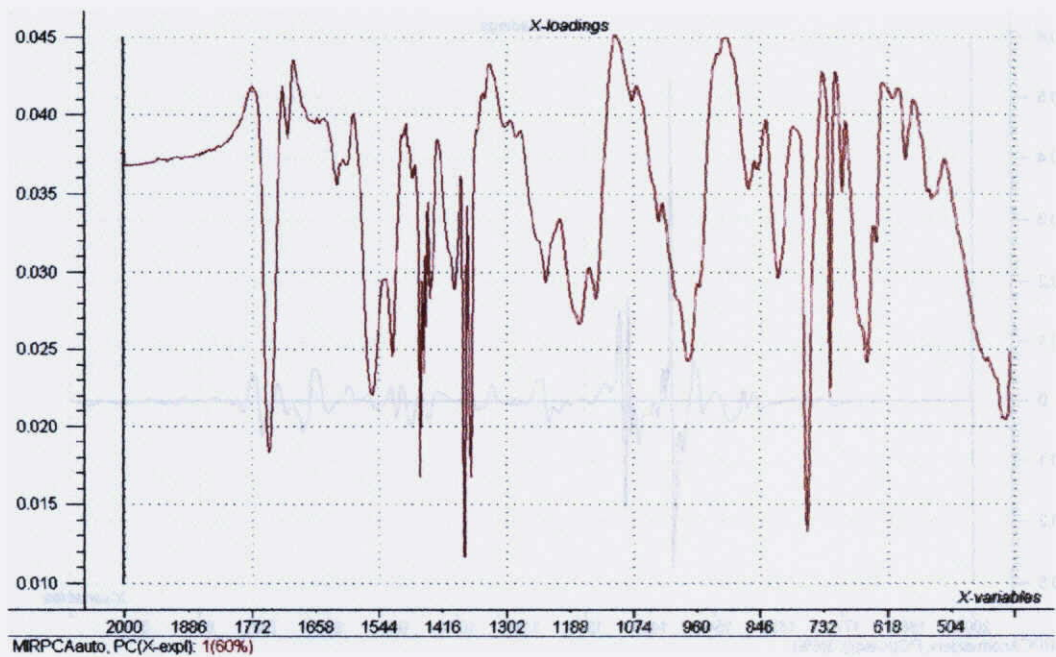


FIG. 6.40 Gráfico dos Pesos versus PC 1 – MIR: espectros autoescalonamento

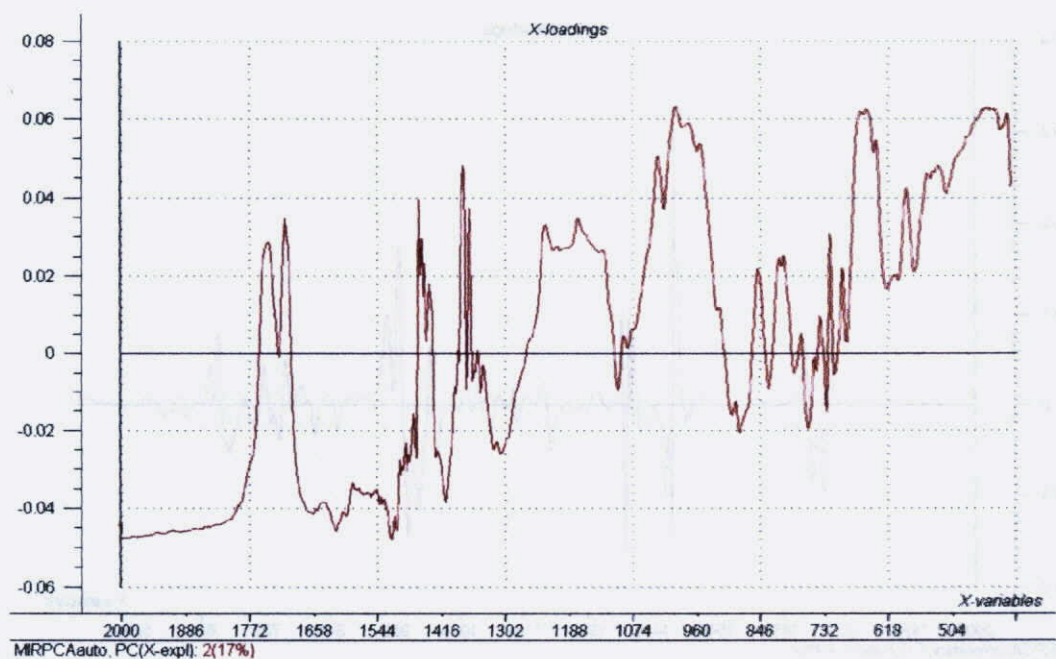


FIG. 6.41 Gráfico dos Pesos versus PC 2 – MIR: espectros autoescalonamento



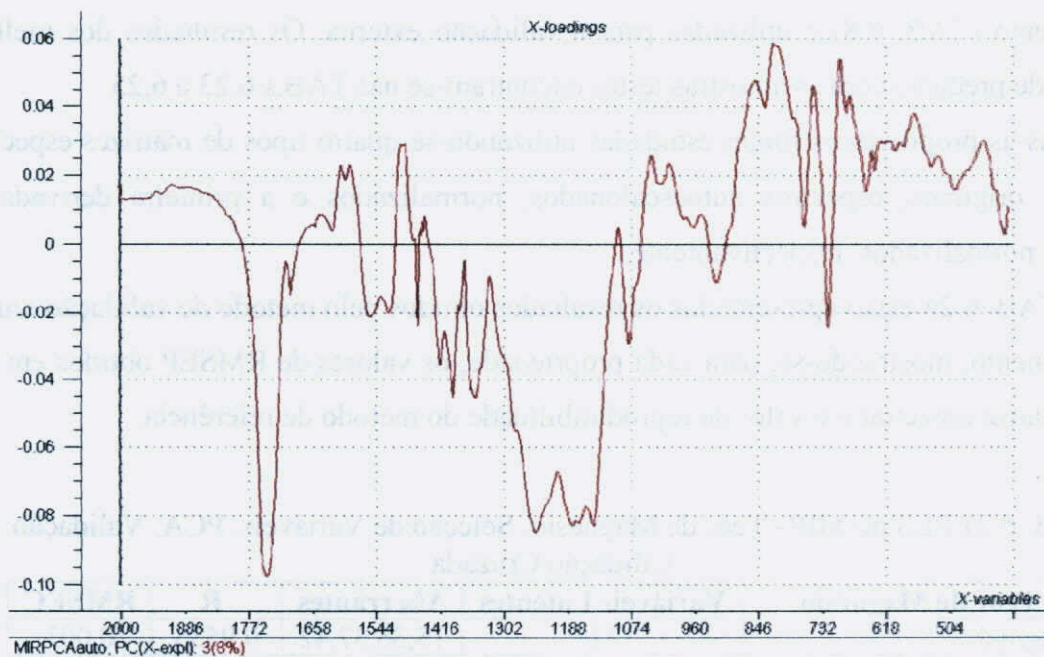


FIG. 6.42 Gráfico dos Pesos versus PC 3 – MIR: espectros autoescalonamento

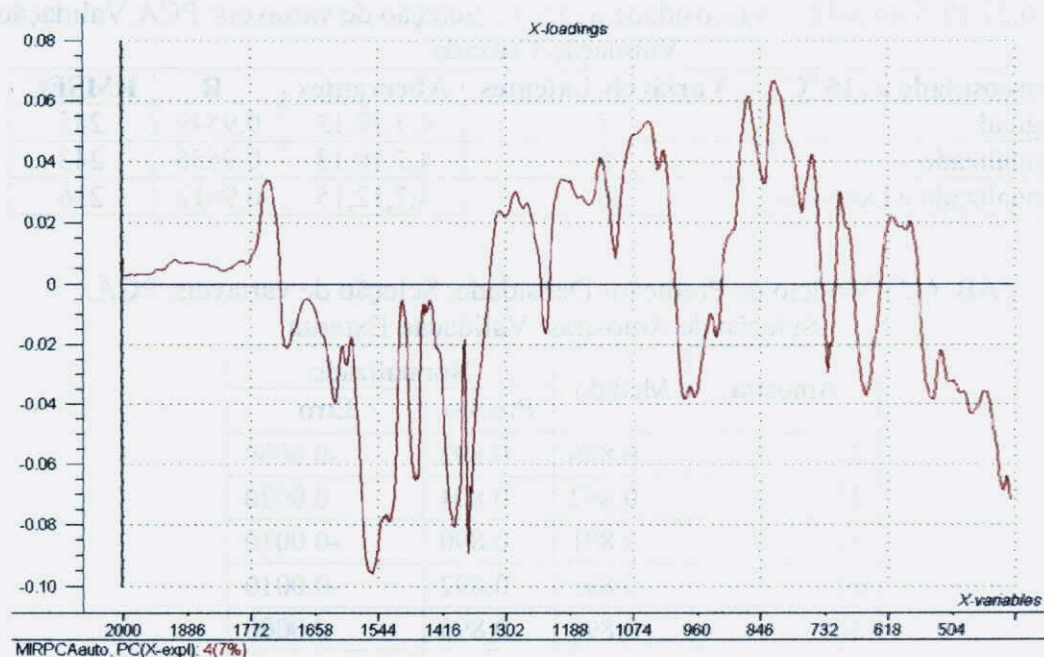


FIG. 6.43 Gráfico dos Pesos versus PC 4 – MIR: espectros autoescalonamento

Nas TAB.s 6.21 e 6.22, encontram-se os resultados da calibração obtidos para as propriedades viscosidade a  $-15\text{ }^{\circ}\text{C}$  e teor de magnésio. Devido à pequena quantidade de amostras (45), optou-se pela validação cruzada como método de validação.

Foram gerados modelos de calibração para as propriedades densidade, viscosidade a  $100^{\circ}\text{C}$  e ponto de fluidez sem a presença das amostras testes, separadas pelo método de

ordenamento (TAB. 6.8) e utilizadas para a validação externa. Os resultados dos melhores modelos de predição com as amostras testes encontram-se nas TAB.s 6.23 a 6.25.

Todas as propriedades foram estudadas utilizando-se quatro tipos de matrizes espectrais: espectros originais, espectros autoescalonados, normalizados e a primeira derivada dos espectros normalizados, respectivamente.

Na TAB. 6.26 estão apresentados os resultados obtidos pelo método de validação cruzada e ordenamento, mostrando-se, para cada propriedade, os valores de RMSEP obtidos em cada tipo de matriz espectral e o valor da reprodutibilidade do método de referência.

TAB. 6.21 PLS no MIR- Teor de Magnésio. Seleção de Variáveis: PCA. Validação: Validação Cruzada.

Teor de Magnésio	Variáveis Latentes	Aberrantes	R	RMSEC
Original	11	15,36,37,42	0,9535	0,003
Normalizado	10	11,20,21	0,9527	0,003
Normalizado e Derivado	10	11,20,32,33	0,9575	0,002

TAB. 6.22 PLS no MIR – Viscosidade a -15 °C. Seleção de variáveis: PCA. Validação: Validação Cruzada.

Viscosidade a -15 °C	Variáveis Latentes	Aberrantes	R	RMSEC
Original	5	4,7,10,13	0,9549	245
Normalizado	5	4,7,10,13	0,9556	243
Normalizado e Derivado	6	4,7,12,15	0,9612	236

TAB. 6.23 Modelo de Predição- Densidade. Seleção de variáveis: PCA. Seleção de Amostras: Validação Externa

Amostra	Medido	Normalizado	
		Predito	Erro
3	0.896	0.893	-0.0030
25	0.892	0.894	0.0020
32	0.891	0.890	-0.0010
34	0.893	0.892	-0.0010
36	0.890	0.890	0.0000
37	0.889	0.889	0.0000
52	0.884	0.888	0.0040
54	0.882	0.888	0.0060
58	0.851	0.851	0.0000
64	0.890	0.886	-0.0040
PRESS			0.0001
RMSEP			0.0028



TAB. 6.24 Modelo de Predição- Ponto de Fluidez. Seleção de variáveis: PCA.  
Seleção de Amostras: Validação Externa

Amostra	Medido	Normalizado	
		Predito	Erro
3	-9.000	-9.521	-0.52
7	-18.000	-12.351	5.65
19	-27.000	-26.021	0.98
23	-24.000	-24.691	-0.69
35	-24.000	-23.966	0.03
40	-27.000	-27.452	-0.45
51	-27.000	-26.973	0.03
52	-33.000	-26.115	6.89
59	-36.000	-39.486	-3.49
69	-30.000	-30.072	-0.07
PRESS			93.39
RMSEP			3.06

TAB. 6.25 Modelo de Predição- Viscosidade a 100 °C. Seleção de variáveis: PCA.  
Seleção de Amostras: Validação Externa

Amostra	Medido	Normalizado e Derivado	
		Predito	Erro
1	14.360	14.930	0.57
3	14.640	14.897	0.26
21	20.770	20.553	-0.22
25	18.980	19.115	0.14
36	18.610	18.917	0.31
41	18.740	18.791	0.05
47	15.350	17.940	2.59
48	14.870	17.811	2.94
52	19.620	18.732	-0.89
61	13.880	14.090	0.21
PRESS			16.74
RMSEP			1.29

TAB. 6.26 Resultados da Validação no MIR. Seleção de variáveis: PCA. Validação:  
Validação Cruzada e Ordenamento

Propriedades	RMSEP			Reproduz.
	Esp. Orig.	Esp. Norm.	Esp. NID.	
Densidade	0,0037 g/cm <sup>3</sup>	0,0028 g/cm <sup>3</sup>	0,0048 g/cm <sup>3</sup>	0,0005g/cm <sup>3</sup>
Viscosidade a 100 °C	1,68 cSt	1,52 cSt	1,29 cSt	0,13 cSt
Viscosidade a -15 °C	319 cP	313 cP	335 cP	670 cP
Ponto de Fluidez	3 °C	3 °C	3 °C	6 °C
Teor de Magnésio	0,005 %(m)	0,005 %(m)	0,004 %(m)	0,032 %(m)

### 6.5.2.2 DISCUSSÃO - ANÁLISE DE COMPONENTES PRINCIPAIS (MIR)

Para a densidade e viscosidade a  $-15\text{ }^{\circ}\text{C}$ , os menores valores de RMSEP foram observados com a matriz dos dados normalizados.

Para o ponto de fluidez, todos os valores de RMSEP foram iguais em todas as matrizes espectrais utilizadas.

Para o teor de magnésio e viscosidade a  $100\text{ }^{\circ}\text{C}$ , os menores valores foram observados com a primeira derivada dos espectros normalizados.

Quando se compara o método de seleção de variáveis PCA com a utilização dos espectros completos sem nenhuma seleção de variáveis pode-se observar que:

- Para a densidade, a utilização da seleção de variáveis PCA gerou um aumento nos valores de RMSEP, enquanto que para a viscosidade a  $100\text{ }^{\circ}\text{C}$ , a utilização da seleção de variáveis PCA causou uma melhora nesses valores para a matriz dos dados originais e normalizados. Ambos os métodos apresentam resultados superiores a duas vezes o valor da reprodutibilidade, podendo ser empregados apenas para controle de processo.
- No caso da viscosidade a  $-15\text{ }^{\circ}\text{C}$ , todas as demais técnicas geraram valores abaixo da reprodutibilidade, logo podem ser usados tanto no controle de processo como em métodos laboratoriais.
- Para o ponto de fluidez, os resultados encontrados para o método de seleção de variáveis PCA e para os espectros completos foram basicamente iguais para todas as técnicas.
- Para o teor de magnésio, houve um pequeno aumento nos valores de RMSEP para as matrizes dos dados originais e normalizados quando se utilizou o método de seleção de variáveis PCA. Apesar disso, ambos os métodos podem ser utilizados tanto em laboratórios como em controle de processo.

TAB. 6.26 Resultados de Validação de MIR. Seleção de variáveis PCA. Validação Cruzada e Ordenamento

Propriedades	RMSEP		
	Eq. Orig.	Eq. Norm.	Eq. 1 <sup>a</sup> Der.
Densidade	0,0023 (g/cm <sup>3</sup> )	0,0028 (g/cm <sup>3</sup> )	0,0018 (g/cm <sup>3</sup> )
Viscosidade a $100\text{ }^{\circ}\text{C}$	1,28 cSt	1,22 cSt	1,28 cSt
Viscosidade a $-15\text{ }^{\circ}\text{C}$	319 cP	313 cP	327 cP
Ponto de Fluidez	$3\text{ }^{\circ}\text{C}$	$3\text{ }^{\circ}\text{C}$	$3\text{ }^{\circ}\text{C}$
Teor de Magnésio	0,002 (g/mg)	0,002 (g/mg)	0,004 (g/mg)



## 6.6 MÉTODO DE SELEÇÃO DE VARIÁVEIS R<sup>2</sup>

Após a geração de uma matriz de correlação **R** para cada propriedade, retira-se a primeira linha, referente à propriedade em questão, e, a partir daí, selecionam-se os números de onda que apresentam as menores correlações entre a variável **X** (variáveis independentes) e o vetor **y** (variáveis dependentes) na primeira coluna da matriz **R**, sendo então retirados da etapa de geração do modelo (TAB.6.27 e TAB.6.28). Os modelos de calibração PLS são gerados com os números de onda restantes, que possuem maior correlação. O objetivo desse método de seleção é verificar a correlação entre a matriz **X** e o vetor **y**, eliminando assim ruídos e fatores não relevantes dos modelos de calibração.

TAB. 6.27 Número de onda retirado do modelo PLS - NIR

PROPRIEDADE – NIR	NÚMERO DE ONDA (nm)
Densidade	1,55-80,105-116,128-180,191-206,225-234,264-357,376-380,388-468,478-505,509-680
Ponto de Fluidez	1-16,50-171,198-218,233-241,274-382,388-501,513-680
Teor de magnésio	1,3-25,51-71,122-212,226-232,235,246-311,319-328,339-355,387-469,479-481,487-501,504-577
Viscosidade a -15°C	1,3,6-8,10-11,53-74,85-90,119-208,226-235,251-309,324-349,388-468,481-680
Viscosidade a 100°C	14-48,69-106,135-154,177-194,241-346,375-377,379-382,384-388,399-456,480-485,491-680

TAB. 6.28 Número de onda retirado do modelo PLS - MIR

PROPRIEDADE – MIR	NÚMERO DE ONDA (cm <sup>-1</sup> )
Densidade	1-216,230-295,298-430,440-442,493-539,609-615,627-638,654-677,714-794
Ponto de Fluidez	110-163,173-247,253-295,302-309,313-330,349-375,382-493,515-544,567-575,606-607,624-655,671-723,728-736
Teor de magnésio	1-41,86-128,134-148,159-191,205-296,304-306,311-312,356-436,464-527,547-550,554-577,589-630,637-653,666-754
Viscosidade a -15°C	1-129,132-182,216-266,274-306,311-315,357-436,490-526,622-640,655-678,697-704,716-794
Viscosidade a 100°C	1-142,152-303,307-310,313-375,439-465,536-586,589-590,594-653,794



### 6.6.1 MÉTODO DE SELEÇÃO DE VARIÁVEIS $R^2$ – NIR

Nas TAB. 6.29 e 6.33, encontram-se os resultados da calibração obtidos para as cinco propriedades estudadas, da mesma forma que na Seção 6.4.1 (TAB.s 6.2 a 6.6).

Na TAB. 6.34, estão apresentados os resultados obtidos pelo método de validação cruzada e ordenamento, mostrando-se, para cada propriedade, os valores de RMSEP obtidos para cada tipo de matriz espectral e o valor da reprodutibilidade do método de referência.

TAB. 6.29 PLS no NIR- Densidade. Seleção de Variáveis:  $r^2$ .  
Validação: Validação Cruzada.

Densidade	Variáveis Latentes	Aberrantes	R	RMSEC
Original	6	27,55	0,9949	0,0012
Autoescalamento	6	27,55	0,9943	0,0013
Normalizado	6	27,55	0,9930	0,0015
Normalizado e Derivado	8	27,55	0,9970	0,0010

TAB. 6.30 PLS no NIR- Viscosidade a 100 °C. Seleção de Variáveis:  $r^2$ .  
Validação: Validação Cruzada.

Viscosidade a 100 °C	Variáveis Latentes	Aberrantes	R	RMSEC
Original	11	9,20,24,31	0,9835	0,47
Autoescalamento	12	1,20,41,45	0,9856	0,43
Normalizado	12	9,20	0,9803	0,50
Normalizado e Derivado	9	10,20,36,45	0,9815	0,49

TAB. 6.31 PLS no NIR- Viscosidade a -15 °C. Seleção de Variáveis:  $r^2$ .  
Validação: Validação Cruzada.

Viscosidade a -15 °C	Variáveis Latentes	Aberrantes	R	RMSEC
Original	12	4,7,12,23	0,9912	115
Autoescalamento	7	4,5,7,12	0,9768	186
Normalizado	8	4,7,12,17	0,9723	192
Normalizado e Derivado	7	4,7,10,25	0,9658	218

TAB. 6.32 PLS no NIR- Ponto de Fluidez. Seleção de Variáveis:  $r^2$ .  
Validação: Validação Cruzada.

Ponto de Fluidez	Variáveis Latentes	Aberrantes	R	RMSEC
Original	9	17,33,45,52	0,9697	1,88
Autoescalamento	7	17,33,45,52	0,9639	2,05
Normalizado	9	17,33,45,52	0,9700	1,87
Normalizado e Derivado	5	6,7,33,45	0,9537	2,26



TAB. 6.33 PLS no NIR- Teor de Magnésio. Seleção de Variáveis:  $r^2$ . Validação: Validação Cruzada.

Teor de Magnésio	Variáveis Latentes	Aberrantes	R	RMSEC
Original	8	10,28,35	0,8699	0,004
Autoescalamiento	6	21,25,30	0,8562	0,004
Normalizado	8	25,28	0,8451	0,004
Normalizado e Derivado	4	-	0,7700	0,005

TAB. 6.34 Resultados da Validação no NIR. Seleção de Variáveis:  $r^2$ . Validação : Validação Cruzada

Propriedades	RMSEP				Reproduz.
	Esp. Orig.	Esp. auto.	Esp. Norm.	Esp. NID.	
Densidade	0,0016 g/cm <sup>3</sup>	0,0016 g/cm <sup>3</sup>	0,0018 g/cm <sup>3</sup>	0,0015 g/cm <sup>3</sup>	0,0005g/cm <sup>3</sup>
Viscosidade a 100 °C	0,72 cSt	0,69 cSt	0,76 cSt	0,71 cSt	0,13 cSt
Viscosidade a -15 °C	268 cP	292 cP	329 cP	351 cP	670 cP
Ponto de Fluidez	3 °C	2 °C	3 °C	3 °C	6 °C
Teor de Magnésio	0,006 %(m)	0,006 %(m)	0,006 %(m)	0,007 %(m)	0,032 %(m)

#### 6.6.1.1 DISCUSSÃO - SELEÇÃO DE VARIÁVEIS $R^2$ (NIR)

Para a densidade, o menor resultado de RMSEP foi observado com a primeira derivada dos espectros normalizados. Contudo, devido aos baixos valores de RMSEP para as demais técnicas empregadas, os resultados podem ser considerados muito bons.

Para a viscosidade a -15 °C, observou-se que o menor resultado foi obtido sem nenhum pré-tratamento. Além disso, o valor de RMSEP para a matriz dos dados autoescalados também se apresentou satisfatório.

Para o ponto de fluidez e viscosidade a 100 °C, os menores valores obtidos para o RMSEP foram observados através do autoescalonamento dos dados.

Para o teor de magnésio, os menores valores obtidos para o RMSEP foram encontrados nos modelos com os dados originais, autoescalados e normalizados.

Comparando-se o método de seleção de variáveis  $r^2$  com os espectros completos e a seleção de variáveis PCA pode-se observar que:

- Densidade: os menores valores de RMSEP obtidos foram adquiridos através da seleção de variáveis  $r^2$ .
- Viscosidade a 100 °C: os dados obtidos com a seleção de variáveis PCA e com o espectro completo apresentaram resultados de RMSEP inferiores àqueles obtidos com o uso da



seleção de variáveis  $r^2$ , com exceção do emprego da normalização no método de seleção de variáveis PCA.

- Viscosidade a -15 °C: a utilização da seleção de variáveis  $r^2$  apresentou uma melhora significativa nos valores de RMSEP para a matriz dos dados originais e autoescalonados empregado no espectro completo. No entanto, para a primeira derivada dos espectros normalizados, o mesmo apresentou um resultado superior a duas vezes o valor da reprodutibilidade.
- Teor de magnésio: os menores resultados de RMSEP foram obtidos com a utilização do espectro completo sem nenhuma seleção de variáveis. Somente na normalização da seleção de variáveis PCA é que o resultado de RMSEP foi igual ao valor para a utilização do espectro completo. Entretanto, mesmo o pior valor obtido de RMSEP é menor que duas vezes o valor da reprodutibilidade do método, ou seja, estes três métodos analisados podem ser utilizados em laboratórios e no controle de processo.
- Ponto de fluidez: somente com os valores de RMSEP para os modelos obtidos com os dados autoescalonados da seleção de variáveis  $r^2$  e com os dados originais dos espectros completos é que os resultados foram inferiores, quando comparados com os demais modelos dos três métodos empregados no presente trabalho. No entanto, com ou sem seleção de variáveis, os modelos podem ser empregados, tanto em controle de qualidade como no controle de processos.

No caso da viscosidade a 100 °C e da densidade, nenhum dos três métodos pode ser empregado em controle analítico. Somente podem ser aplicados no controle de processos, que tolera uma margem maior de erros.

#### 6.6.2 MÉTODO DE SELEÇÃO DE VARIÁVEIS $R^2$ – MIR

Nas TAB.s 6.35 e 6.36, encontram-se os resultados da calibração obtidos para as propriedades viscosidade a -15°C e teor de magnésio. Também foram gerados modelos de calibração para as propriedades densidade, viscosidade a 100°C e ponto de fluidez sem a presença das amostras testes (TAB. 6.8). Estes modelos estão descritos nas TABs. 6.37 a 6.39, utilizando-se os quatro tipos de matrizes espectrais, como descrito anteriormente na seção 6.4.2 (TAB.s 6.9 a 6.13).



Na TAB. 6.40, estão apresentados os resultados obtidos pelo método de validação cruzada e ordenamento, mostrando-se, para cada propriedade, os valores de RMSEP obtidos para cada tipo de matriz espectral e o valor da reprodutibilidade do método de referência.

TAB. 6.35 PLS no MIR- Teor de Magnésio. Seleção de Variáveis:  $r^2$ . Validação: Validação Cruzada.

Teor de Magnésio	Variáveis Latentes	Aberrantes	R	RMSEC
Original	9	11,20,36	0,9299	0,003
Autoescalamento	8	6,11,20,36	0,9415	0,003
Normalizado	12	7,11,20,33	0,9506	0,002
Normalizado e Derivado	10	11,20,21,33	0,9636	0,002

TAB. 6.36 PLS no MIR- Viscosidade a -15 °C. Seleção de Variáveis:  $r^2$ . Validação: Validação Cruzada.

Viscosidade a -15 °C	Variáveis Latentes	Aberrantes	R	RMSEC
Original	5	4,7,12,13	0,9600	237
Autoescalamento	7	4,7,12,44	0,9666	216
Normalizado	6	4,7,12,13	0,9638	226
Normalizado e Derivado	9	4,7,12,27	0,9851	148

TAB. 6.37 Modelo de Predição- Densidade – Seleção de Variáveis:  $r^2$   
Seleção de amostras: validação externa

Amostra	Medido	Normalizado e Derivado	
		Predito	Erro
3	0.896	0.892	-0.0040
25	0.892	0.893	0.0010
32	0.891	0.893	0.0020
34	0.893	0.892	-0.0010
36	0.890	0.890	0.0000
37	0.889	0.889	0.0000
52	0.884	0.886	0.0020
54	0.882	0.889	0.0070
58	0.851	0.852	0.0010
64	0.890	0.893	0.0030
PRESS			0.0001
RMSEP			0.0029

TAB. 6.38 Modelo de Predição- Ponto de Fluidez – Seleção de Variáveis:  $r^2$   
Seleção de amostras: validação externa

Amostra	Medido	Original	
		Predito	Erro
3	-9.000	-13.125	-4.13
7	-18.000	-13.225	4.78
19	-27.000	-23.568	3.43
23	-24.000	-26.167	-2.17
35	-24.000	-23.838	0.16
40	-27.000	-27.376	-0.38
51	-27.000	-26.279	0.72
52	-33.000	-24.781	8.22
59	-36.000	-37.946	-1.95
69	-30.000	-29.179	0.82
PRESS			128.99
RMSEP			3.59

TAB. 6.39 Modelo de Predição- Viscosidade a 100°C – Seleção de Variáveis:  $r^2$   
Seleção de amostras: validação externa

Amostra	Medido	Autoescalamento	
		Predito	Erro
1	14.360	14.577	0.22
3	14.640	14.977	0.34
21	20.770	17.966	-2.80
25	18.980	18.808	-0.17
36	18.610	19.504	0.89
41	18.740	18.532	-0.21
47	15.350	15.909	0.56
48	14.870	15.637	0.77
52	19.620	17.328	-2.29
61	13.880	13.800	-0.08
PRESS			15.06
RMSEP			1.23

TAB. 6.40 Resultados da Validação no MIR. Seleção de Variáveis:  $r^2$ . Validação : Validação Cruzada e Ordenamento

Propriedades	RMSEP				Reproduz.
	Esp. Orig.	Esp. auto.	Esp. Norm.	Esp. NID.	
Densidade	0,0032 g/cm <sup>3</sup>	0,0031 g/cm <sup>3</sup>	0,0036 g/cm <sup>3</sup>	0,0029 g/cm <sup>3</sup>	0,0005g/cm <sup>3</sup>
Viscosidade a 100 °C	1,24 cSt	1.23 cSt	1,28 cSt	2,84 cSt	0,13 cSt
Viscosidade a -15 °C	316 cP	305 cP	300 cP	257 cP	670,7cP
Ponto de Fluidez	4 °C	4 °C	4 °C	5 °C	6 °C
Teor de Magnésio	0,005 %(m)	0,005 %(m)	0,005 %(m)	0,005 %(m)	0,032 %(m)



### 6.6.2.1 DISCUSSÃO - SELEÇÃO DE VARIÁVEIS $R^2$ (MIR)

Para a densidade e viscosidade a  $-15\text{ }^{\circ}\text{C}$ , os menores valores de RMSEP foram observados com a matriz dos dados derivados após a normalização.

Para o teor de magnésio, todas as técnicas empregadas apresentaram basicamente o mesmo resultado.

Para o ponto de fluidez, a utilização da primeira derivada dos espectros normalizados levou a um maior valor de RMSEP, enquanto que as demais técnicas apresentaram valores iguais de RMSEP.

Para viscosidade a  $100\text{ }^{\circ}\text{C}$ , o menor valor de RMSEP foi adquirido após o autoescalonamento dos dados.

Comparando-se o método de seleção de variáveis PCA e os espectros completos sem nenhuma seleção de variáveis com a seleção de variáveis  $r^2$ , pode-se observar que:

- Densidade: com a utilização da seleção de variáveis PCA e  $r^2$ , ocorre um aumento nos valores de RMSEP. Em todos os métodos, os resultados apresentados são superiores a duas vezes o valor da reprodutibilidade, portanto, só podem ser empregados para controle de processo.
- Viscosidade a  $100\text{ }^{\circ}\text{C}$ : o menor valor de RMSEP foi obtido sem a utilização de seleção de variáveis para os dados derivados após normalização. Mesmo assim, esse valor é superior a duas vezes a reprodutibilidade, só podendo ser usado para controle de processo.
- Viscosidade a  $-15\text{ }^{\circ}\text{C}$ : os menores valores foram obtidos para a matriz dos dados originais sem nenhuma seleção de variáveis do espectro completo e com o emprego da primeira derivada dos espectros normalizados para o método de seleção de variáveis  $r^2$ . Entretanto, no método do espectro completo, a utilização da matriz dos dados autoescalados só pode ser usada no controle de processo, devido aos seus valores, que são superiores a 2 vezes o valor da reprodutibilidade.
- Ponto de Fluidez: os menores valores de RMSEP foram obtidos com a utilização do método de seleção de variáveis PCA e espectro completo (com exceção dos dados autoescalados). Para a seleção de variáveis  $r^2$  e para os dados autoescalados do espectro completo, os resultados foram superiores a duas vezes o valor da reprodutibilidade.

- Teor de magnésio: os resultados encontrados são muito parecidos, podendo-se fazer uso de todos os três métodos tanto em controles analíticos como em controle de processos.

## 6.7 RESUMO DOS MELHORES RESULTADOS

De forma geral, os resultados obtidos no presente trabalho podem ser resumidos nas TAB. 6.41 e 6.42., onde estão apresentados os resultados mais expressivos da validação no MIR e no NIR.

TAB. 6.41 Resultados da Validação no NIR

Propriedades	Pré-tratamento	Método	RMSEP	Reproduz.
Densidade	Derivação após Normalização	$r^2$	0,0015 g/cm <sup>3</sup>	0,0005g/cm <sup>3</sup>
Viscosidade a 100 °C	original	PCA	0,61 cSt	0,13 cSt
Viscosidade a -15 °C	original	$r^2$	268 cP	670 cP
Ponto de Fluidez	Autoescalonamento e original	Espectro Completo e $r^2$	2 °C	6 °C
Teor de Magnésio	Normalização e original	Espectro Completo e PCA	0,005 %(m)	0,032 %(m)

TAB. 6.42 Resultados da Validação no MIR

Propriedades	Pré-tratamento	Método	RMSEP	Reproduz.
Densidade	Normalização	Espectro Completo	0,0024 g/cm <sup>3</sup>	0,0005g/cm <sup>3</sup>
Viscosidade a 100 °C	Derivação após Normalização	Espectro Completo	0,78 cSt	0,13 cSt
Viscosidade a -15 °C	Original	Espectro Completo	256 cP	670 cP
Ponto de Fluidez	Todos os Pré-tratamentos utilizados	Espectro Completo e PCA	3 °C	6 °C
Teor de Magnésio	Derivação após Normalização (PCA), normalização (PCA) e autoescalonamento	Espectro Completo e PCA	0,004%(m)	0,032 %(m)



Para a validação na espectroscopia do infravermelho próximo, o método de seleção de variáveis  $r^2$  apresentou o menor valor de RMSEP para a densidade empregando a primeira derivada dos espectros normalizados. Isto se deve possivelmente à correção dos efeitos multiplicativos da normalização (como aqueles gerados por pequenas variações no percurso óptico) associada à eliminação dos desvios da linha base de um conjunto de espectros, que ocorre na primeira derivada dos espectros normalizados.

Também para a viscosidade a  $-15\text{ }^\circ\text{C}$ , a validação na espectroscopia do infravermelho próximo teve o menor valor de RMSEP obtido com o método de seleção de variáveis  $r^2$  para a matriz dos dados originais, demonstrando que, para essa propriedade, o impacto das variações aleatórias (ruídos, deslocamento de comprimentos de onda, etc.) e não uniformes não interferem de modo significativo no resultado.

Para o ponto de fluidez, os menores valores foram obtidos com os dados autoescalados no método de seleção de variáveis  $r^2$  e com os dados originais dos espectros completos. Isso se dá provavelmente devido ao autoescalamento aplicar a técnica do escalamento pela variância, que resulta em variáveis com pesos equivalentes, atribuindo-se a mesma importância para todas as variáveis, independente da sua dimensão, e à eliminação dos ruídos através da centralização dos dados na média. No caso da matriz dos dados originais, o impacto das variações aleatórias e não uniformes não interferem de modo significativo no resultado.

Para a viscosidade a  $100\text{ }^\circ\text{C}$ , o melhor valor foi obtido com o emprego da seleção de variáveis PCA para a matriz dos dados originais.

No caso do teor de magnésio, os menores valores foram obtidos com a matriz dos dados normalizados no método de seleção de variáveis PCA e com os dados originais para o espectro completo.

Para a densidade, o melhor resultado de RMSEP encontrado na espectroscopia no MIR foi obtido com o emprego da normalização como pré-tratamento para o espectro completo.

No caso da viscosidade a  $100\text{ }^\circ\text{C}$ , para o MIR, a utilização da primeira derivada dos espectros normalizados como pré-tratamento para o espectro completo apresentou o menor resultado de RMSEP.

No MIR, o ponto de fluidez de todas as matrizes espectrais empregadas para a seleção de variáveis PCA, com exceção da matriz dos dados autoescalados do espectro completo, apresentaram o mesmo bom resultado de RMSEP.

Para a viscosidade a  $-15\text{ }^{\circ}\text{C}$ , o menor resultado encontrado no presente trabalho, para a espectroscopia no MIR, foi obtido com a matriz dos dados originais para o espectro completo.

Para o teor de magnésio, os resultados obtidos para o espectro completo empregando os dados originais, normalizados e derivados após normalização e para a seleção de variáveis PCA com o emprego dos dados derivados após normalização, foram os melhores resultados encontrados.



## 7.0 CONCLUSÕES

Este trabalho mostra a eficiência das técnicas de espectroscopia no infravermelho próximo (NIR) e médio (MIR) para a construção de modelos de propriedades físico-químicas de óleos lubrificantes comerciais. Associou-se a estas técnicas, o método de regressão multivariada por mínimos quadrados parciais (PLS), empregando-se pré-tratamentos dos dados originais e diferentes métodos de seleção de variáveis.

A análise de componentes principais, realizadas nas 71 amostras, permitiu a separação das diversas amostras de maneira satisfatória para as propriedades viscosidade a 100°C, ponto de fluidez e densidade, tanto para os espectros NIR como MIR. Os diversos grupos, provenientes de indústrias diferentes puderam ser separados de acordo com suas classes. Além disso, para um mesmo fabricante, alguns tipos de óleos, como o A<sub>1</sub> e o A<sub>2</sub>, puderam ser agrupados. Entretanto, os óleos lubrificantes semi-sintéticos, por serem constituídos por misturas de óleos minerais e sintéticos, aparecem com propriedades muito semelhantes à dos óleos minerais, não se separando totalmente em um agrupamento.

Com relação à espectroscopia no NIR aplicada aos óleos lubrificantes comerciais, verificou-se, após as etapas de calibração e validação, que essa técnica pode ser utilizada no controle de processos industriais para as propriedades físico-químicas estudadas. Isso é possível porque os mesmos toleram certa margem de erro, ou seja, não necessitam da mesma exatidão fornecida pelos métodos de referência. Além disso, para o ponto de fluidez, teor de magnésio e viscosidade a -15 °C, as técnicas também poderiam ser aplicadas como método de controle de qualidade, já que foi possível obter resultados de RMSEP inferior a duas vezes o valor do método de referência. Deve-se destacar que a exatidão fornecida pelos métodos de referência para a densidade e viscosidade é bastante restritiva, sendo praticamente inatingível.

De acordo com os resultados apresentados, observou-se que os pré-tratamentos com melhor desempenho para os modelos no NIR foram o autoescalamento, seguido pela normalização.

Nos resultados apresentados, pode-se observar que os melhores modelos construídos para densidade e viscosidade a -15 °C foram obtidos com o método de seleção de variáveis  $r^2$ .

Nos casos da viscosidade a 100 °C, o menor valor de RMSEP foi conseguido com o emprego da seleção de variáveis PCA.

Para o teor de magnésio, os menores valores de RMSEP foram obtidos com o método de seleção de variáveis PCA e com o espectro completo.

Já para o ponto de fluidez, os menores valores de RMSEP foram obtidos com o método de seleção de variáveis  $r^2$  e com o espectro completo.

O emprego da espectroscopia no MIR, aplicada aos óleos lubrificantes comerciais, apresentou resultados semelhantes ao da espectroscopia no NIR, ou seja, resultados superiores (menores valores de RMSEP) para as propriedades teor de magnésio e viscosidade a  $-15^{\circ}\text{C}$ , e inferiores (maiores valores de RMSEP) em relação às demais propriedades.

Para a densidade e viscosidade a  $100^{\circ}\text{C}$ , pode-se observar que os melhores modelos de predição foram obtidos sem nenhuma seleção de variáveis.

No caso da viscosidade a  $-15^{\circ}\text{C}$ , os melhores modelos de predição foram obtidos sem nenhuma seleção de variáveis e com a seleção de variáveis  $r^2$ .

Para o ponto de fluidez, os modelos obtidos com a seleção de variáveis  $r^2$  não foram satisfatórios em relação aos demais métodos, podendo ser empregados apenas em controle de processos.

No caso do teor de magnésio, os resultados encontrados foram todos muito parecidos, podendo ser empregados tanto em controles analíticos como para controle de processo.

Os pré-tratamentos que apresentaram melhores resultados para os modelos no MIR foram a normalização e a primeira derivada dos espectros normalizados.



## 8.0 REFERÊNCIAS BIBLIOGRÁFICAS

- ALBUQUERQUE, Jackson da Silva; **Desenvolvimento de um Sensor Óptico para a determinação de Hidrocarbonetos Aromáticos em águas Empregando a Espectroscopia no Infravermelho Próximo (NIR)**. Dissertação - UFPE
- ANDRADE, J. M.; SANCHEZ, M. S.; SARABIA, L. A.; *Chem. Intell. Lab. Syst* 46, 41(1999).
- ARAÚJO, M.A.S.; **Óleos Lubrificantes – Introdução, Ensaio de Caracterização e Óleos Básicos** – Universidade Corporativa (2000).
- BARROS, A.S.; RUTLEDGE, D.N.; *Chem. Intell. Lab. Syst* . 70, 99 (2004).
- BARROS, A.S.; RUTLEDGE, D.N.; *Chem. Intell. Lab. Syst* . 73, 245 (2004).
- BEEBE, Kent; Pell, J.; Seasholtz, M.B. **Chemometrics: a practical guide**. New York: Wiley, 1999.
- BORIN, Alessandra.; **Aplicação de Quimiometria e Espectroscopia no Infravermelho no Controle de Qualidade de Lubrificantes**. Dissertação- UNICAMP. 2003.
- CERQUEIRA, E.O.; ANDRADE, J.C.; POPPI, R.J.; *Química Nova*, 24-6 , 864 (2001).
- CHUNG, H.; KU, Min-Sik; LEE, Joon-Sik; *Vibrat. Spectrosc.* **20**, 155 (1999).
- FILHO, P.A.C; POPPI, R.J.; *Química Nova* **25-1**,46 (2002).
- FILHO, Paulo Augusto da costa; **Estudos Comparativo entre Técnicas de Inteligência Artificial e Modelos Lineares em Determinações Quantitativas no Infravermelho Próximo**. Dissertação- UNICAMP. 2003.
- ESTIENNE, F.; DESPACHE, F.; WALCZAK, B.; NOORD O.E; MASSART, D.L.; *Chem. Intell. Lab. Syst* . **73**, 207 (2004).
- FARAH, Marco Antônio; **Apostila de “ Caracterização de frações de petróleo – Parte 2 – Lubrificantes** – CEN-SUD – 1987.
- FERRÃO, M.F; CARVALHO, C.W.; MULLER, E.I; DAVANZO, C.U; *Ciên. Technol. Aliment.* 24-3, **864**, 2001.
- FERREIRA, M.; ANTUNES, A.; MELGO, M.; VOLPE, P. *Quimimetria I: Calibração Multivariada, um tutorial*. *Química Nova*, São Paulo, v. 22, n. 7, 1999.
- LIMA, F.S.G; ARAÚJO, M.A.S; BORGES, L.E.P; *Tribol. Int.* **36**, 691 (2003).

- LIMA, Flavio Santos de Gusmão; **Espectroscopia no infravermelho próximo aplicada a óleos lubrificantes básicos**. Dissertação- IME. 2003.
- LIMA, F.S.G; ARAÚJO, M.A.S; BORGES, L.E.P; *J. Near Infrared Spectrosc.* **12**, 159 (2005).
- MORGANO, M.A.; MORIYA, C.; FERREIRA, M.M.C.; *Brazilian Journal of Food Technology*, **6**, 77, 2003.
- MORGANO, M.A.; FARIA, C.G.; FERRÃO, M.F.; BRAGAGNOLO, N.; FERREIRA, M.M.C.; *Ciê. Tecnol. Aliment. Technology*, 25-1, **25**, 2005.
- MOURA, CRS; CARRETEIRO, R.P.; *Lubrificantes e Lubrificação*; Makron Books, **31**, (1998).
- MONTGOMERY, D.C; RUNGER, G.C; *Estatística Aplicada E Probabilidade Para Engenheiros*. 2ª Edição, Editora: LTC - Livros Técnicos e Científicos, 2003.
- PASCHOAL, J.; BARBOZA, F. D.; POPPI, R. J.; *J. Near Infrared Spectrosc.* **11**, 211 (2003).
- OTTO, Matthias; **Chemometrics – Statistics and Computer Application in Analytical Chemistry**. New York: Wiley. 1999
- PASQUINI, C. *J. Braz. Chem. Soc.*, **14**, 198 (2003).
- PEDRO, A. M. K.; **Determinação Simultânea e Não- Destrutiva de Sólidos Totais e Solúveis, Licopeno e Beta- Caroteno em Produtos de Tomate por Espectroscopia no Infravermelho Próximo Utilizando Calibração Multivariada**. Dissertação- UNICAMP. 2004.
- RIBEIRO, Fabiana Alves de Lima; **Aplicação de Métodos de Análise Multivariada no Estudo de Hidrocarbonetos Policíclicos Aromáticos**. Dissertação- UNICAMP. 2003.
- ROCHA, Eduardo Azevedo Lourenço; **Análise em Gasolina por Espectroscopia no Infravermelho Através de Métodos Quimiométricos**. Dissertação- IME. 2002.
- RUNGE, P.R.F; DUARTE, G.N; *Lubrificantes nas Industrias*: Editora TRIBOCONCEPT , 1989.
- SCAFI, Sérgio Henrique Frasson.; **Espectroscopia no Infravermelho Próximo para identificação de medicamentos Falsificados**. Tese de Mestrado –UNICAMP-2000.
- SENA, M.M.; COLLINS, K.E.; *Química Nova*, 24 - 3, 331 (2001).
- SILVERSTEIN, R.M.; BASSLER, G.C.; MORRIL, T.C.; **Identificação Espectrométrica de Compostos Orgânicos**, Editora - Guanabara dois (1998).



SIMÃO, Cristina A. **Avaliação das Propriedades de Óleos Diesel Provenientes De Diferentes Refinarias Através De Espectros NIR Utilizando Técnicas Quimiométricas.** Dissertação – IME. 1999.

SKOOG; HOLLER; NIEMAN; **Princípios de Análise Instrumental**, 5ª Edição, Editora ARTMED (2002).

WOLD, S.; SJOSTROM, M.; *Chem. Intell. Lab. Syst.* **44**, 3 (1998).

WOLD, S.; SJOSTROM, M.; ERIKSSON, L.; *Chem. Intell. Lab. Syst.* **58**, 109 (2001).

UNSCRAMBLER users guide, version 6.11, software package for multivariate calibration and prediction. CAMO-ASA. 1998.

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)



[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)